1

AGARD CONFERENCE PROCEEDINGS No. 43

AGARD CP No. 43

AD 692540

AGARD
ADVISORY GROUP FOR AEROSPACE RESEARCH & DEVELOPMENT

7 RUE ANCELLE 92 NEUILLY SUR SEINE FRANCE

Inertial Navigation—Systems and Components

★

MAY 1968

NORTH ATLANTIC TREATY ORGANIZATION

INITIAL DISTRIBUTION IS LIMITED
FOR ADDITIONAL COPIES SEE BACK COVER

AGARD Conference Proceedings No. 43

NORTH ATLANTIC TREATY ORGANIZATION

ADVISORY GROUP FOR AEROSPACE RESEARCH AND DEVELOPMENT

(ORGANISATION DU TRAITE DE l'ATLANTIQUE NORD)

INERTIAL NAVIGATION -
SYSTEMS AND COMPONENTS

SEP 11 1969

Papers presented at two Meetings of the Guidance and Control Panel of AGARD,
in Oxford, England during 1967 and at Braunschweig, Germany, in May 1968

PAGES ______

ARE

MISSING

IN

ORIGINAL

DOCUMENT

681.2.082.16:629.7.058.82:
527.6:061.3

Printed by Technical Editing and Reproduction Ltd
Harford House, 7-9 Charlotte St. London, W.1.

FOREWORD

In September 1966, the then Ad Hoc Guidance and Control Panel of AGARD (Advisory Group for Aerospace Research and Development of the North Atlantic Treaty Organisation) held its first symposium in Paris on the subject of man machine interactions in guidance and control systems. In March 1967, the Panel held its second symposium, also in Paris, on the subject of reliability aspects of guidance and control systems. By the end of the second symposium the significance and growth potential of this new Panel activity of AGARD was rather clearly established and this was amplified by the large attendance and great interest shown by attendees at the next two Panel symposia. The first of these was held in Oxford, England, in September 1967, and its subject was "Inertial Navigation: Systems and Applications". The second was held in Braunschweig, Germany, in May 1968, and its subject was "Inertial Navigation: Components". During this same time the Guidance and Control Panel was taken off ad hoc status and given full fledged status as one of the AGARD Panels.

This volume of Symposia Proceedings of the Oxford and Braunschweig meetings represents the first publication of the Guidance and Control Panel. The reader will find contained herein coverage of the broad areas of inertial navigation: components, systems and applications. Fourteen papers were presented at the Oxford Symposium and all but two of these are included here as indicated in the table of contents. Twenty-one papers were presented at the Braunschweig Symposium and all but three of these are included here, also as indicated in the table of contents.

It is a pleasure to express gratitude to a number of people who made these Panel Symposia Proceedings possible. First of all, a great deal of credit must be given to Professor W.Wrigley, who in his tireless efforts as the first Chairman of the Guidance and Control Panel, was the person most responsible for establishing it as a full fledged AGARD Panel. These published Proceedings are, of course, one of the fruits of his efforts.

Professor B.Fraeijs de Veubeke, as the first Panel Vice Chairman, and Mr H.G.R.Robinson, the present Panel Vice Chairman, deserve a great deal of credit also for the development of the Panel's programs and activities. To Mr H.G.R.Robinson, Chairman of the Oxford Symposium, and to Leonard Sugerman, Colonel USAF, Chairman of the Braunschweig Symposium, goes the credit for developing the programs of these symposia. A great deal of credit must also be given to Mr Frank Sullivan for his important contributions to Panel activities and to the development of these Proceedings. Appreciation is also expressed to Professor G.Whitfield and to Monsieur l'Ingénieur en Chef G.Bonnevale, who in their capacity as Panel Associate Editors, provided valued assistance in the development of these Panel Proceedings. Finally, gratitude is also expressed to the British Government for offering to serve as the host country for the Oxford Symposium and to the Government of the Federal Republic of Germany for hosting the Braunschweig Symposium. The facilities in both instances were excellent and contributed significantly to the success of both Symposia.

C.T.LEONDES
Editor
Guidance & Control Panel, AGARD

CONTENTS

Page

FOREWORD iii

INTRODUCTION:
INERTIAL PRINCIPLES IN NAVIGATION AND GUIDANCE
by C.S.Draper vii

PART I: Systems and Applications

TECHNIQUES AND PHILOSOPHY OF MIXING OR AIDING INERTIAL NAVIGATION WITH OTHER NAVIGATION AIDS
by G.E.Roberts 1

NAVIGATION PAR INERTIE SUR DE LONGUES DISTANCES
par B. de Crémiers 15

COST-OF-OWNERSHIP PHILOSOPHY APPLIED TO INERTIAL NAVIGATION SYSTEMS
by W.J.Laubendorfer, R.V.Plank and E.J.DeNezza 31

A STUDY OF STRAPDOWN INERTIAL NAVIGATION AT THE MARSHALL SPACE FLIGHT CENTER
by F.P.Daniel, G.B.Doane III and R.R.Kissel 49

DESCRIPTION OF A STRAPDOWN INERTIAL MEASUREMENT UNIT
by Joseph Yamron 71

FIELD TEST, ALIGNMENT AND TRIMMING OF THE ELDO-A PLATFORM
by Manfred Pütz 97

PREDICTION DE L'ERREUR D'UNE CENTRALE INERTIELLE
by Miguel da Silveira 109

FACTORS INFLUENCING THE CHOICE OF COMPUTER FOR AN INERTIAL NAVIGATION SYSTEM
by P.Wilson 119

OPTIMAL USE OF REDUNDANT DATA IN INERTIAL NAVIGATION SYSTEMS
by F.G.Unger 127

THE IMPACT OF STATISTICAL ESTIMATION ON INERTIAL NAVIGATION
by Larry D.Brock 147

APPLICATION OF INERTIAL TECHNOLOGY TO AIRBORNE GRAVIMETRY
by Elmer J.Frey and Raymond B.Harlan 167

PRESENT-DAY NORTH-FINDING SYSTEMS AND THEIR FUNDAMENTAL LIMITATIONS
by R.T.Trayner 191

Page

PART II: Components

SYSTEM PARAMETERS AS CONSIDERED IN THE DESIGN OF A SMALL INERTIAL NAVIGATION COMPUTER
by J.Kydd 207

THE APPLICATION OF STATISTICAL ESTIMATION TECHNIQUES TO INERTIAL COMPONENT CALIBRATION
by George T.Schmidt 231

EULER ANGLE STRAPPED-DOWN COMPUTER
by Alan van Bronkhorst 251

A CONCEPT OF DIGITAL AVIONICS
by J.F.Bussell 281

MICROPLASTICITY IN METALS FOR PRECISION INSTRUMENTS
by O.H.Wyatt 297

GYRO BALL BEARINGS - TECHNOLOGY TODAY
by Albert P.Freeman 313

CONICAL GAS BEARINGS FOR GYROSCOPE SPIN AXIS SUPPORT
by D.Faddy and T.L.Ellis 357

STRAPDOWN GUIDANCE COMPONENT RESEARCH
by Richard J.Hayes 375

THE GYROFLEX GYROSCOPE
by Richard F.Cimera and Michael Napolitano 401

THE ROTOR-VIBRAGYRO
by E.Mühlenfeld 411

THE OSCILLOGYRO
by R.Whalley and D.W.Alford 419

THE PIPA (PULSED INTEGRATING PENDULOUS ACCELEROMETER)
by George J.Bukow 429

L'ACCELEROMETRE ONERA A GRANDE SENSIBILITE
par Michel Delattre 459

THE VIBRATING STRING ACCELEROMETER
by Robert O.Bock 475

TECHNICAL REPORT ON THE QUARTZ RESONATOR DIGITAL ACCELEROMETER
by Norman R.Serra 487

AN OPTICALLY PUMPED NUCLEAR MAGNETIC RESONANCE GYROSCOPE
by James H.Simpson and I.A.Greenwood 517

GYROMETRES LASER ET APPLICATIONS
par J.M.Catherin et B.Dessus 525

STATUS AND FUTURE OF FLUID COMPONENTS
by D.L.Wright and W.G.Wing 537

INTRODUCTION

INERTIAL PRINCIPLES IN NAVIGATION AND GUIDANCE

by

C.S.Draper

Massachusetts Institute of Technology, Cambridge,
Massachusetts, USA

INERTIAL PRINCIPLES IN NAVIGATION AND GUIDANCE

C. S. Draper

1. INTRODUCTION

Navigation is the process of purposefully determining motion so that desired objectives are achieved. Activity of this kind has been among the important endeavors of man for the many thousands of years he has traveled on foot, with animals, with sledges, with wagons, in canoes, in boats and in sailing vessels. Modern times have intensified the need for always higher performance from navigation as the technology of transportation has expanded to include power driven ships, subsonic and supersonic airplanes, helicopters, VTOLs, orbiting capsules and soon-to-come craft for travel between the earth and the moon with perhaps the planets ahead within a few decades.

Moving entities incorporating means for changing speed and direction are essential for accomplishing desired results by reaching selected destinations. The function of navigation is to determine the modifications in motion needed to cause the entity involved to follow a path that results in mission success. The action pattern that must be carried through in order to realize these results may be considered in terms of five functions:

1. Selection of desired results and the generation of procedures for their achievement through plans and programs which give the necessary position, speed, direction, etc., as functions of time in a convenient reference space.

2. Determination of the actual situation at known instants of real time.

3. Comparison of actual situations as they develop in real time with desired situations at corresponding programmed instants to determine deviations of actual states from desired states of the essential geometrical quantities.

4. Computation from the deviation data of corrections needed to bring the actual states into agreement with the desired states.

5. Application of these corrections as commands for the control system of the moving entity.

When the last four functions are carried out continuously and the control system operates effectively, the result is a course for the moving entity that carries it over an actual path that closely approaches the programmed path with respect to the reference space. Until fairly recent years all of the five component functions were carried out by one or more men. At first, a single man provided all the component functions from planning to making proper changes in the directions in which he may have been walking when it became obvious that corrections were needed. Much later developments led to operators working levers and steering ships in response to instructions from superior officers responsible for navigation, while today many systems are automatic and only require human attention for monitoring purposes. The second, third and fourth functions, which are determination of position, indication of deviations and computation of corrections belong to the method of traditional NAVIGATION. When the functions of planning and programming and of control to carry out corrections are added to navigation, the resulting complex of functions is GUIDANCE.

Before either navigation or guidance can be effective, the moving entity must be STABILIZED so that it is capable of responding properly to control commands. In effect, stabilization is the process of causing all the essential variables associated with the moving entity to remain near desirable equilibrium states established either naturally or by instrumental means. CONTROL acts by changing these equilibrium states so that in seeking the new stabilization conditions the moving entity makes corrections toward desired positions and states of motion.

It is the purpose here to review the fundamental functional requirements of STABILIZATION, CONTROL, NAVIGATION and GUIDANCE with particular attention to the essential services that are provided for modern equipments by devices based on inertial principles. After some twenty-five years of practice and discussion, inertial systems are still sometimes regarded as basically rather low performance devices struggling in competition with optical, radio and radar navigation equipments. This is an unfortunate opinion which has retarded progress and substantially increased the cost of advancing modern navigation technology. This technology will inevitably achieve performance and reliability increased significantly beyond the levels now realized in practice. These improvements will come from engineering developments of inertial components and subsystems integrated into overall systems incorporating radiation contacts of various kinds with computers to couple all subsystems into complexes of always improving effectiveness. The technical details involved provide subject matter for the presentations of this conference. This particular paper is not concerned with particular equipments, rather its purpose is to "set the stage" for later papers by reviewing the generalized requirements of navigation and the contributions that can be provided by devices based on inertial principles.

2. GEOMETRICAL REQUIREMENTS

Missions for moving entities must be defined in geometrical terms with time as the basic "fourth dimension" that ties all other variables together in a continuous sequence of instants uniquely associated with the general progress of changes in the essential physical quantities. The formulation of plans and programs to specify the chain of events needed to achieve a desired result from motion of a controlled entity must be referred to some geometrical reference space within which destinations and moving entities can be described in position and orientation. For example, if the moving entity is an airplane and the destination is a certain city, it is convenient to select a reference space determined by the earth coordinates of latitude, longitude, north and the vertical. The position of a destination city is easily read from a map in terms of its position in this reference space. It is the problem of navigation to locate the airplane in this same space and relate its position and motion to points on the path it must follow in reaching its goal.

With known landmarks visible in good weather, simple sightings make it possible to pinpoint an airplane, to select a suitable path for reaching a selected destination and to determine the actual path and the necessary corrections. When terrestrial landmarks are not available but star sightings are possible, the traditional method of using celestial space for an auxiliary geometrical reference, correcting for earth rotation by chronometer time and applying known stellar positions, make it possible to determine aircraft positions by the traditional computations of celestial navigation. When both terrestrial and celestial contacts by visual observations are eliminated by unfavorable circumstances, man-generated radiation with known patterns determined by cooperative ground stations make it possible to navigate by means of ground and airborne receivers and properly designed computing aids. With visibility denied and cooperative radiation stations lacking, self-contained on-board methods must be used to measure or predict changes in vehicle position as functions of time. The time honored method of dead-reckoning based on estimates of air speed, flight directions, wind speed, wind direction, etc., is always available for the computation of future positions, but cannot provide high accuracy.

Navigation and guidance subsystems operating without outside cooperation during relatively long periods of time and free of dependence on estimates of air speed, flight direction or wind can give much better results than conventional dead reckoning. These systems developed only during recent years depend upon computations using data sensed by on-board instruments mechanically held in known orientations with respect to earth coordinates. Equipment of this kind is made possible by inertial principles which are inherently ideal and free from saturation effects within the necessary operating ranges for components and subsystems of navigation and guidance to operate for considerable periods of time without external contacts for geometrical information.

Indications from inertial sensors are not absolute, but represent changes from artificially introduced initial conditions that are related in known ways to time and the selected geometrical reference space. Once initial conditions are set, inertial equipments operate by a modern form of dead reckoning that uses, not estimates of speed and direction, but measurements of velocity changes along coordinate axes associated with instrumentally maintained internal coordinates. Position changes with respect to the external reference space are computed from the indicated velocity change components. It is obvious that for systems of this kind the accuracy of results depends on the performance of the inertial sensors, their associated computers and other necessary subsystems. Many theoretical, engineering and technological problems are important for the technology of inertial systems that are beyond the scope of this paper, which is concerned only with an introduction to general principles.

3. COMPLETE NAVIGATION AND GUIDANCE SYSTEMS

Navigation must have accurate contacts with a geometrical reference space in which the desired motions of the moving entity are completely defined. For paths involving terrestrial problems, earth coordinates are particularly well suited for reference purposes and are used by direct visual contacts by artificial long wave radiation, by use of intermediate celestial objects and by instrumented inertial reference members. The essential function of the geometrical reference space is to provide an orthogonal coordinate system to which displacements of the moving vehicle may be related in terms of components associated with the space in which the vehicle path is described. When ground stations are available these components may be measured by optical, radio or radar means, so there is no question of coordinate relationships, the coordinates actually used in measurement are identical with those to which vehicle motions are referred. Radiation receivers, either on the ground or in the vehicle, produce signals suitable for processing by properly designed computing subsystems which generate outputs that lead to indications of position, velocity, off-course deviations, velocity correction commands etc., that are the useful outputs of navigation systems. To be effective these outputs must be used as the inputs for a control system that actually causes the vehicle motion to change in ways that force it to follow a path that leads to the desired destination.

When the functions of supplying destination information and programs for operation, of providing reference coordinates, sensing motion components, computing control commands and executing these commands are integrated together, the resulting complex is a GUIDANCE SYSTEM. If the program function and the control function are omitted the result is a NAVIGATION SYSTEM.

CONTROL SYSTEMS serve to interface navigation systems with the vehicles that they guide, in the sense that they receive information level correction commands as inputs and provide changes in vehicle motion as their outputs. In order for this action to be effective, the control system must first provide STABILIZATION which, in effect, is control about equilibrium conditions corresponding to "no-change" input commands. For example, if an airplane is to properly execute commands related to earth coordinates it must have orientational information supplied to its controls so that right-turn, left-turn, up-down and roll-right, roll-left are related to the on-board geometrical reference in the same way that these angular motions are related to the flight path defined in terms of the external reference coordinates.

It has been common practice for almost forty years to provide aircraft with stabilization and control reference coordinates by means of gyroscopic instruments. The bank and climb indication, which in effect provides an instrumentally established horizontal plane reference with accuracies in the range of a few degrees, is as satisfactory today for flight control purposes as it was when General Doolittle first used it for blind flight in 1929. In a similar way the gyroscopic turn indicator has shown, and shows today, right and left turns of aircraft with about the same level of accuracy.

The combination of these two gyroscopic instruments to form automatic pilots has been, and still is, effective for the purposes of aircraft stabilization and control, but devices for sensing changes in linear motion have in the past never been used in autopilots to provide indications of navigational information. A fundamental limitation of conventional gyroscopic aircraft instruments lies in their inherent indication inaccuracies of a few degrees and their inherent drift rates of approximately ten degrees per hour. When it is recalled that one degree angle between local gravitational directions corresponds to sixty miles distance on the earth's surface, it becomes obvious that ordinary aircraft instruments can not provide suitable internal reference coordinates for navigation and guidance systems with accuracy requirements of one mile error for each hour of operation after an initial fix.

4. INERTIAL GUIDANCE

Inertial guidance is the term applied to the operation of an equipment subsystem which has three generalized functions. The first function is that of establishing and maintaining an internal geometrical reference space which has an associated set of coordinates and means by which their orientations may be accurately adjusted to the desired relationships with the external reference space in which the destination to be reached and the path to be followed are specified. The second function of inertial guidance is to receive specific force, which is the vector resultant of gravity force and inertial reaction force, in terms of components related to the internal coordinate system and to generate output signals that accurately represent these components. The third subsystem function is computation, in recent times almost universally carried out in digital terms, by equipment which accepts these output signals for inputs and, as its own outputs, provides control commands and the indications needed for effective vehicle operation. Control, a fourth function which implements control commands, has its low power level input side interfaced with the computer and its output side driving a control subsystem that is usually closely associated with the vehicle itself. When men are involved, displays, input pushbuttons, and operating knobs to select and adjust modes of operation are included in guidance systems for the purposes of communication and interaction between human operators and the arrangements that cooperate with them as inanimate working partners.

The technology of computing equipment is now capable of meeting all the requirements of inertial guidance for storing data on plans and programs, for carrying out calculations and for transmitting information to proper interface points. Similarly, the technology of displays and relay systems is so well developed that almost any desired arrangement may be realized by which a man can be informed by systems as to actual states of affairs and be given the power to impose his will upon the machine as to its modes of operations, its tasks and the disposition of its results. Control systems capable of forcing actual vehicle motion into substantial agreement with programmed motion are well developed and generally designed into operational vehicles.

With good design and effective quality control, computing, display and control subsystems of inertial guidance equipment will not limit performance of overall systems. Reliability is another matter, in that it is sometimes determined by these three functions but is more often set by electronics, either associated with the reference member or with the computer and the displays. This state of affairs means that currently the overall system performance levels are generally established by the inertial sensors. For this reason these components become subjects for particular attention in any discussion of inertial guidance systems.

5. INERTIAL SENSORS

Inertial sensors must provide functions of two kinds. The first function of an individual unit is that of sensing small angular deviations with respect to inertial space about a sharply defined input axis and producing an electrical output signal, with its sense determined by the direction of the deviation and its magnitude proportional to the size of the deviation. In practice three angular deviation sensors of this kind are used in a triad configuration with their input axes held accurately at right angles on the geometrical reference member to which the instruments are rigidly mounted.

The reference member has three degrees of angular freedom with respect to a base upon which it is mounted and which in turn is installed in the vehicle to be guided. The three axes corresponding to the degrees of freedom are each fitted with a tightly operating servo-drive that draws its input signals from the properly combined signals from the angular deviation sensors. In operation the overall servo-drive system keeps a set of orthogonal axes fixed to the reference member accurately aligned with the orientations of the three sensors for which their output signals have null (i.e. minimum magnitude) levels. Current servo practice is developed to levels that allow the realization of drives capable of keeping the reference member coordinate axes in substantially perfect alignment with the signal-null orientations of the sensors. This is accomplished by designing high torque capabilities into the servos and providing "noise free" output signals from the angular deviation sensors. This state of affairs means that the reference member departs from ideal performance only because of sensor imperfections that cause the orientations for which their output signals have their nulls to differ from initial settings. Uncertainties associated with "drift" effects of this kind are the limiting factors for angular deviation sensor performance.

The second type of sensor required for inertial guidance system operation is the receiver for specific force, which is the vector resultant of gravitational force and the reaction forces that accompany accelerations of material particles with respect to inertial space. If their operation is to be ideal these sensors must respond only to the specific force component along a single sharply defined direction, which for any given instrument is its input axis. In practice, three specific force sensors are mounted rigidly on the inertial reference member with their input axes mutually at right angles and carefully fixed in known orientations with respect to the geometrical reference member. With this arrangement the sensor output signals in the process of being generated are already accurately resolved into components associated in a known way with the coordinates of the geometrical reference member. Because of practical considerations, specific force sensors are usually designed to generate output signals representing, not specific force directly, but the first integral of specific force starting from initial conditions selected for convenience in the operations currently in progress.

Performance for inertial sensors ranges over several orders of magnitude in terms of inputs and the quality of the corresponding outputs. Details of construction belong to the subject matter discussed in the technical papers which follow, not to this introductory paper.

6. PLACE OF INERTIAL PRINCIPLES IN NAVIGATION AND GUIDANCE

Navigation and guidance deal with problems of dynamic geometry with changes in orientation and in position, as they change with time, as the elements of basic interest.

To deal effectively with the situations of navigation it is necessary to establish reference for time, for orientation, and for position. With such reference available navigation and guidance are matters of determining changes in time and correlating changes in orientation and position with these time changes. The accuracy of navigational results depends on the performance of the means for measuring the basic physical quantities. The effectiveness of the means employed in navigation depends upon its ability to work under operating conditions ranging from completely free and continuous interchange of information with the environment to situations in which this interchange is not available for long periods of time in which moving vehicles must be located and guided.

It has been noted that, when visual, radio and radar contacts with the ground and the sky can be used, so that the earth and the stars may serve directly as references, long established methods are available for navigation. When continuous radiation contacts are eliminated, self-contained equipment able to provide high quality navigation for extended periods of time is required. Today navigation must deal with motion in three dimensions over a very wide range of speeds, but the general situation is somewhat similar to that which existed for sailing ships before the day of John Harrison, 1735, who invented the first operational chronometer. The problem was that of longitude, which could be solved only by an accurate means for indicating time so that the rotation of the earth could be properly taken into account for the interpretation of celestial observations. Once the marine chronometer was available navigation became a process of greatly improved accuracy and reliability.

Harrison faced the need for a timekeeper to operate for several months, there being no radio stations to provide updating signals, within accuracy limits unheard of in the year 1713, when the British Government offered various rewards for solutions to the problem of longitude. He accepted the basic law of physics that a unique relationship exists everywhere between the acceleration imparted to any massive particle and the force associated with this acceleration. He understood that gravity interfered with this relationship in a way that centuries later would be stated by Einstein in his PRINCIPLE OF EQUIVALENCE, but he also realized that there were ways of eliminating this disturbing effect. Relatively had not as yet been discovered and fortunately introduced no significant effects in Harrison's time, or in ours of today, because as yet vehicle speeds are very small compared to the velocity of light.

Realizing that inertial reaction forces were available everywhere without the need for mechanical or visual contacts with outside spaces, Harrison made use of effects that were being successfully used in clocks and watches after millennia of applying considerably more awkward and less accurate methods. He refined the mechanism by which the elastic action of a spring, working against the inertial force reaction of a miniature flywheel carried by low friction bearings, produced an oscillatory mechanism with a period adjustable to a desired constant value which, with proper temperature control and leveling by proper gimbals, produced the timekeeping accuracy required for good longitude measurements.

It is history that, after Harrison showed the way during the middle 1700's, many chronometers were built with performance that removed time measurements from the unenviable position of limiting the accuracy of navigation. It is pertinent, but not basically important, that radio stations with crystal-controlled oscillators, atomic clocks and very precise determinations of star transits now provide excellent monitoring for on-board electronic clocks, which certainly eliminates inaccuracies in time as a limiting factor in the overall processes of navigation and guidance.

The marine chronometer and the many devices that are its descendants have solved the availability of accurate time indications on board moving vehicles. However, navigation, which to be completely satisfactory must solve problems in dynamic geometry by self-contained on-board equipment of high accuracy and reliability, requires that internal reference coordinates and internally generated indications of vehicle position, velocity and acceleration be continuously available within moving vehicles. Monitoring by radiation contacts of any and all kinds is certainly desirable but not necessary during extended time periods and is to be considered as a functional part of overall systems, not as an arrangement competing with inertial principles.

Sensing of angular deviations with respect to inertial space with self-contained instruments is accomplished with angular momentum provided by gyroscopic rotors. Forced precession of such rotors away from an initial orientation requires torque, and that can be used to drive a viscous integrator which produces an angular input to a signal generator, whose output represents the angular deviation of the gyro unit case with respect to its reference orientation. The gyro output signal from a single gyro provides one of the three inputs needed to establish and hold an inertial reference member in its reference orientation and is maintained within limits dependent upon gyro drift rates which are made as low as possible by design and manufacture.

Sensing of linear displacements of components of the reference member from an initial position in inertial space depends upon an instrument that uses an unbalanced mass to impose a torque on a calibrated restraint element. Signals dependent upon the output of this calibrated element are proportional to the input specific force component along the sensor input axis. Effective double integration of this acceleration signal and correction of the result for the effects of gravity produces indications of changes in position.

Inertial components have been in existence for some time to prove the feasibility of inertial navigation and guidance. The future surely holds advances in technology that will lead to ever-increasing usefulness of navigation and guidance equipment. Details of these future developments are the basic subject matter of the papers that follow.

7. CONCLUSIONS

Navigation and guidance are dependent on time, reference coordinates in which the desired path of the guided vehicle is defined and a means for measuring vehicle displacements with respect to these coordinates. Time is generally available everywhere from watches, clocks, chronometers, vibrating crystals, tuning forks, molecular drivers, earth's rotation against the stars etc., and certainly does not present more than routine engineering problems today.

Geometrical orientation for reference coordinates may be taken from lines of sight to stars of known location on the celestial sphere, from the earth, from instrumentally driven inertial reference members, and from any other available space that may be convenient. Just as instruments for measuring time based on vibrations determined by the dynamic balance between elastic forces and inertial reaction forces are used because of their continuous operation with no more than infrequent monitoring contacts, inertia reaction effects mechanized by gyro instruments are essential for the accurate sensing of angular deviations from initially set orientations.

Once an instrumentally established inertial reference member is available with means for close alignment to an external reference space, sensors for specific force along well defined input axes can provide signals that accurately indicate the components of specific force, the resultant of gravity and inertia reaction force, along the three orthogonal inertial reference member axes. Proper computational processing of these signals, to correct for gravitational effects and to transform results into a convenient external reference space, gives navigational information on position, velocity, direction and ground speed for the moving vehicle. This information can be displayed to human operators and may also be used as signals in automatic control.

Inertial principles make it possible to have continuous outputs for navigation and guidance from on-board equipment for significant periods of time without monitoring by contacts with external space. Operation of this kind is essential when radiation contacts are not available and is very helpful when such contacts are intermittent. The usefulness of basic inertial equipment obviously depends upon the accuracy of its indications as functions of time, its reliability in operation and its hours of satisfactory performance per dollar of combined first price and total costs of maintenance. In any case, inertial equipment is here to stay - it does not represent gadgetry that will disappear after its novelty has worn off. The false claims of particular mechanisms will in time be replaced by true engineering information and sound technology.

Many implementations of inertial sensors and all the other components of complete inertial systems have been advertised and reduced to practice. It is not the purpose of this paper to discuss these details or the sort of performance that has been achieved in operational systems. However, one remark based on much personal experience may be in order. This remark is that the quality and overall hours performance per dollar of total cost will surely be much better for future systems than it is for present systems.

PART I

TECHNIQUES AND PHILOSOPHY OF MIXING OR AIDING INERTIAL NAVIGATION WITH OTHER NAVIGATION AIDS

by

G.F.Roberts

Royal Aircraft Establishment,
Farnborough, Hampshire, UK

SUMMARY

A broad treatment is given of the requirement for the use of hybrid Inertial Navigation systems in the light of continual development in the capability of pure Inertial systems, and the availability of new navigation aids and computing techniques. The main characteristics of the pure Inertial system are described, with particular emphasis on those aspects of performance and integrity which require the support of other aids, and the advantages and limitations of some early examples of hybrid systems are discussed.

Future trends in hybrid inertial navigation systems are discussed, emphasising the important role that modern digital computing techniques are going to have in such systems for both military and civil airborne use.

NOTATION

R	local earth's radius of curvature
w	effective vertical gyro drift rate
V_I	aircraft's velocity component as measured by Inertia
V_D	aircraft's velocity component as measured by Doppler
ΔV	error in measurement of aircraft's velocity component
Ω	earth's angular rotation
Λ	aircraft latitude
$\Delta\psi$	angle between platform azimuth datum and true North
$\Delta\psi_\epsilon$	error in measurement of $\Delta\psi$
K_1, K_2	scaling factors in feedback loops
ΔD	error in resolved Doppler speed component

TECHNIQUES AND PHILOSOPHY OF MIXING OR AIDING INERTIAL NAVIGATION WITH OTHER NAVIGATION AIDS

G.E. Roberts

1. INTRODUCTION

We may define a hybrid Inertial Navigation (I.N.) system as a system in which the overall capability is based on the use of I.N. in conjunction with one or more other navigation aids of a basically different type. The term overall capability is here used in its broadest possible sense to include accuracy, operational flexibility, error distribution characteristics, reliability, and system credibility. System performance is defined here as including the first two of these characteristics and system integrity the last three, and the objective of hybrid systems can then be defined as optimising system performance and integrity.

This paper examines the case for "hybrid" as opposed to "pure" I.N. systems in the airborne application in the light of continual development in I.N. performance, and the availability of new navigation aids and computing techniques. The term Mixed I.N. system is normally employed for a hybrid system in which the accent is on accuracy of the combined system. Since the required accuracy is only attained with both sensors operative, this type of hybrid system does not in general add much to system integrity. In "Aided" systems the additional aid is used to improve system integrity or operational flexibility. The tendency towards ever increasing I.N. accuracy will obviously be accompanied by a tendency towards the use of Aided I.N. rather than Mixed I.N. systems.

As far as the additional navigation aids are concerned we shall be talking about two main types of aid, according to whether the prime information is defined in aircraft or in ground axes. The first type are of the Dead Reckoning type, typified by Doppler Radar, and in general they provide continuous information which has been very suitable for use with mixed I.N. systems using analogue techniques. The second group are normally called "position-fixing" aids and have tended in the past to be available only at discrete intervals and have been used mainly in simple aided I.N. systems. The use of modern digital computing techniques has extended the potential use of discrete fixes and this, coupled with the availability of continuous position information from ground based radio aids and automatic astro, may mean a considerable extension to the use of such aids in the future.

2. I.N. CHARACTERISTICS

It would be pertinent at this stage to examine the performance and integrity characteristics of the pure I.N. system so as to highlight those areas where hybrid systems may be of value. For the sake of simplicity the following treatment of performance ignores the very long term (nominally 24-hour period) oscillatory effects which tend to bound a large proportion of the I.N. errors for a slowly moving vehicle. These effects are in any case drastically modified at aircraft speeds and tend to be of less importance for the relatively short sortie duration of aircraft. It will further be assumed that the Inertial Platform is nominally aligned to local North and East axes in order to simplify the presentation; the general pattern of errors applies equally for other azimuth precession configurations.

Fundamentally I.N. performance depends upon three basic assumptions, each of which can be associated with a particular error source.

The first assumption is that the system continually defines the earth's local vertical axis. In a pure I.N. system the vertical defining system is an undamped Schuler tuned system, and velocity and tilt errors introduced either during initial set up or during take-off and flight are perpetuated in the system as oscillatory terms of approximately 84-minute period. The importance of this type of error term is that it can have a significant effect on the velocity or short term dead-reckoning capability of the system, which in turn can be important in weapon aiming applications and possibly in terminal area navigation for civil airlines.

Next, we have the assumption that the system rate of rotation of the inertial axes are a true measure of the rotation of local earth axes in inertial space. This gives rise to a long term system velocity error of the form Rw, where R is the local earth's radius and w is the precession error term, which includes both the drift rates of the gyros and scale factor errors in the precession terms. While this term contributes to the short term performance its most important effect is in the long term accumulation of position error in the system.

The third basic assumption is that the rate of rotation of the local vertical axis in inertial space can be related to the motion of the aircraft over a rotating earth, which brings in the important factor of azimuth alignment. There are two separate cases to distinguish here:

(i) Where the requirement is to determine the aircraft's velocity in aircraft axes as in some weapon aiming applications, or in hybrid systems with Navigation aids operating in these axes. This gives rise to a long term velocity error term $R\Omega \cos \lambda \Delta\psi$ in the nominally North-orientated inertial axis. Ω is the angular velocity of the earth, λ the latitude, and $\Delta\psi$ the error in alignment of the North axis.

(ii) For normal navigation purposes, or for hybrid systems with navigation aids operating in earth axes, the I.N. must produce outputs defined relative to earth axes. In this case we have an additional long term velocity error in the across-track direction given by $V\Delta\psi$, where V is the velocity of the aircraft, and at very high aircraft speeds this will obviously become an important factor.

In a pure I.N. system the normal method of aligning the system in azimuth is by the technique of gyro-compassing before take off, and relying on the azimuth gyro to maintain this alignment subsequently. Maximum use is made of the fact that the system is practically stationary on the earth during ground gyro-compassing, to speed up the response of the system and to damp out oscillatory terms. When the system has settled down the velocity error in the nominally North channel is given by

$$\Delta V = Rw + R\Omega \cos\lambda \, \Delta\psi .$$

The platform can be precessed in azimuth until ΔV is zero, or the value of ΔV can be used as a measure of the correction to be applied in azimuth. In either case the final error in alignment is given by

$$\Delta\psi_e = \frac{w}{\Omega \cos\lambda} .$$

When operating at latitudes below 70° this means that the azimuth error is reduced by gyro-compassing to a level consistent with other errors in the system.

There is, however, some conflict between the military requirement for rapid reaction time and the alignment time before take off required by gyro-compassing, and so far there are operational or technical reservations on all the schemes designed to overcome this difficulty. When this is taken in conjunction with the fact that there will still be an accumulated error due to drift rate of the azimuth gyro in flight, it is clear that there are great potential benefits to be gained from the possibility of air alignment of the I.N.

In summary we can therefore say that the possible areas where I.N. performance may need enhancement are as follows:

(i) Reducing the effect of the 84-minute oscillatory terms.

(ii) Improving the performance limitation imposed by the drift rate of the vertical gyros.

(iii) The possibility of air alignment.

In the early days of I.N. development the vertical drift rates were well below the requirement for a self-contained navigation system and the accent was on mixed I.N. systems, particularly Doppler Inertia Mixing. At the present stage, with pure I.N. performance capability of the order of one nautical mile per hour or better, the accent is more on Aided Hybrid I.N. systems, with the stress perhaps more on areas (i) and (iii), depending upon the requirement. With the expected line of development in the future, and particularly through the use of the very promising cluster rotation technique in I.N., which considerably reduces the effects in areas (i) and (ii), it could well be that air alignment will be the area in which hybrid I.N. systems will play the most important part in the future from the performance aspect.

The integrity aspects of I.N. are receiving a considerable amount of attention at present because of its application to civil aircraft, both as an attitude reference for flight control and as a navigation aid. From the reliability standpoint the most stringent requirement is set by the flight control application and there is little doubt that this can be met by a duplex or triplex I.N. installation. As far as navigation is concerned the critical aspect is meeting the lateral separation safety standards for civil aircraft; in this context two areas are now highlighted.

The first is the phenomenon known as error distribution tail, a typical example of which is shown in Figure 1. The important feature here is that the frequency of occurrence of large errors is considerably greater than one would expect from a Gaussian distribution, and it is therefore dangerous to predict the chance of a large error on the basis of the standard deviation error. The particular aspect of I.N. which makes it vulnerable from this point of view is that partial malfunctions which can lead to unsafe navigation errors are difficult to detect as such by the crew. Until I.N. development leads to an adequate self-checking capability, there will be obvious advantages in using fixing aids as credibility checks.

The second critical area is concerned with the possibility of common navigation errors in multiplex I.N. systems, produced either by a common aircraft system failure, such as electrical power supply, or by human blunder. The advantage of a hybrid system in this context is that the possibility of common error is fundamentally remote, whereas "designing out" this possibility in a multiplex system can be a formidable task.

3. EXAMPLES OF HYBRID SYSTEMS

The combination of I.N. with Doppler Radar provides an interesting example of hybrid I.N. systems using analogue computing techniques. In the following examples the Doppler Radar velocity output is first resolved through the platform heading angle to provide the components of Doppler derived velocity in inertial axes. This is then subtracted from the inertial velocity and the velocity difference signal has the important property of separating the effects of platform tilt from those of vehicle acceleration, so that it can be used either as a damping term or for modifying the system oscillation frequency as desired.

The simplest form of Doppler-I.N. hybrid system is the Doppler damped system shown in Figure 2. Here the only difference from the pure inertial configuration is the feedback term $K_1(V_I - V_D)$, in which the value of K_1 is adjusted to provide semi-critical damping of the system Schuler tuned oscillations. The system characteristics are practically identical with those of the pure I.N. system, except for the fact that the Schuler tuned oscillations are damped out, with a consequent improvement in the short term velocity performance. The system does, in fact, have an in-flight alignment capability, but this is too slow to be of practical importance, and the same ground alignment techniques as for pure I.N. are normally used with such systems.

Figure 3 shows the configuration for the Doppler-Inertia mixed system in which the objective is to combine the long term navigation accuracy of Doppler with the smooth velocity response of I.N. and have a system capability for in-flight alignment. The velocity difference signal $(V_I - V_D)$ is now fed back along two paths; $K_1(V_I - V_D)$ to the input of the integrator and $K_2(V_I - V_D)$ as an additional vertical gyro precession term. The value of $(K_2 + 1/R)$ determines the overall system period and K_2 is chosen as a compromise between system response time and total velocity error.

Once the value of K_2 is fixed, K_1 is adjusted to provide the required damping and is normally chosen to give approximately semi-critical damping. A much improved overall performance can be obtained if the values of K_2 and K_1 can be varied as the system settles down, but this degree of complexity is probably better handled by the digital computing techniques to be mentioned later.

Considered solely as a system for defining the aircraft's velocity in inertial or aircraft axes, the requirement on inertial accuracy is not particularly severe and the system characteristics can be summarised as follows:

(a) It will settle down to a definition of the vertical good to a few minutes of arc in flight, in a period of the order of ten minutes.

(b) Inertial errors and oscillatory terms in velocity will decay in the same time.

(c) The white noise content of Doppler is considerably reduced by an effective filter equivalent to the system period and damping. There is, however, no equivalent lag in acceleration response which remains practically as good as a pure I.N. system.

(d) The long term accuracy is determined mainly by Doppler.

The steady-state value of $V_I - V_D$ for the nominally North axis is given by the expression

$$(V_I - V_D)_{NS} = \frac{\Delta D}{K_2R + 1} + \frac{R(w + \Omega \cos\lambda\,\Delta\psi)}{K_2R + 1} .$$

ΔD in this expression is the error in the resolved Doppler component and does not include errors in azimuth alignment to true North. The value of $(K_2R + 1)$ is typically about 100 so that the contribution of Doppler + Inertial errors to the right-hand side is very small.

When we consider the Navigation requirement, however, there remains the important problem of alignment in azimuth, and here we have to distinguish between operation with low quality or high quality Inertia. When the vertical gyro wander rate is too large for gyro-compassing to be possible, air alignment in azimuth is only possible by using an additional aid such as astro or ground fixes. Note that the use of ground fixes in this case is not a gyro-compassing mode but a comparison of Doppler Dead Reckoning with fix information and is available world wide.

With high quality inertial components the system is capable of a self-contained azimuth alignment mode in flight by gyro-compassing. This is obtained by using the velocity difference signal in the nominally North channel to precess the azimuth gyro in the direction to bring ΔV to a minimum.

The steady-state value of ψ is then given by

$$\psi_\epsilon = \frac{\dot{\Delta}D + Rw}{R\Omega \cos\Lambda} .$$

Apart from the Doppler error term $\dot{\Delta}D$, this is the same expression as is obtained for ground gyro-compassing, and the effect of vertical gyro wander w and latitude Λ are similar. The effect of the term $\dot{\Delta}D$ is to give an additional error in azimuth which, for subsonic speeds and at normal latitudes, is consistent with the remaining system performance. For very high aircraft speeds or at high latitudes, better performance can be obtained by reverting to the Fix Monitored Azimuth technique.

It will be noted that, if K_2 is reduced to zero and the value of K_1 correspondingly adjusted, the Doppler Mixed system reverts to Doppler Damping and, once the system has settled down from its initial errors, there would be some advantage, in terms of total system error, in making this change. However, it must be realised that, when the gyro-compassing loop is included, the value of $V_I - V_D$ for the North channel is reduced to zero and the system output on the channel will therefore be controlled by the error in Doppler rather than by Inertia, as in the simple damped system.

The performance enhancement obtained in hybrid Doppler-I.N. systems is to some extent at the expense of system integrity. In a Doppler Mixed system using low quality inertial components there is obviously very little capability if the Doppler or the Inertial elements fail. With high quality inertia there is a system capability in the event of Doppler failure, provided this failure is detected and the Doppler cut out early enough. Since catastrophic errors can be introduced into the I.N. in the case of undetected Doppler failure, the integrity of the overall system rather hinges on the availability of a reliable automatic cut-out for Doppler.

The use of ground fixes as a credibility aid for I.N. has already been mentioned, as has the use of fixes in conjunction with Doppler in the Fix Monitored Azimuth technique. The use of fixes for credibility can also be readily extended to the up-dating of I.N. position and this form of hybrid I.N. system is almost invariably used in military aircraft in the target area.

More extensive mixing of I.N. with intermittent ground fixing aids to improve the system velocity performance are not in general amenable to analogue techniques. However, in the particular case in which the I.N. system has been set up on the ground but there has been no time for gyro-compassing, it can be assumed that azimuth alignment is the predominant error and a form of in-flight gyro-compassing using ground fixes can be employed. It is of interest to note that in this case the gyro-compassing is, in terms of the total angular rotation rate, composed of the earth's rotation and the aircraft's motion around the earth. This means that there is no specific latitude limitation to this form of gyro-compassing, but the system is inoperative whenever the aircraft's motion is equal and opposite to the earth's speed at any particular latitude.

One other specialised form of hybrid I.N. system is perhaps worthy of mention at this time. This is the combination of automatic star tracking with the Inertia Navigator in the Astro-Inertial system. With a line of sight defined on two known stars the astro component supplies all the information required to provide a continuous fixing aid and to determine true North, if it is supplied with a knowledge of the local vertical. The role of inertia in the system is to supply the definition of the local vertical and to provide an inertial memory for the periods when the stars are obscured. Since a precise azimuth datum is available when operating in the astro mode and information is also available for correcting the gyro wander rates, the inertial memory will be operating under optimum conditions during the inertial memory mode.

In the Astro-Inertial system the only basic inertial error remaining arises out of the Schuler tuned oscillations of the inertial platform. With the addition of Doppler to the system these can be damped out in the manner indicated previously and the Doppler damped Astro-Inertial system is capable of giving a bounded error of less than one nautical mile for any flight duration and independent of position over the earth. At the same time the system will provide accurate datum directions, both in the local vertical and in azimuth, for flight control purposes.

The main disadvantages of astro in the past have been the fact that the stars can be obscured by clouds and the cost and complexity of achieving the capability of detecting stars by day. However, the modern trend for flying faster and higher are in its favour and it may well be that the cost can be justified in terms of its capability for future aircraft.

4. FUTURE TRENDS

The development of the airborne digital computer is now well advanced and there is little doubt that it will find increasing use for data processing tasks in future aircraft, both military and civil. As applied to the integration of hybrid systems, the flexibility and computing accuracy available with these techniques can make a profound change to the philosophy of mixing in several respects:

(a) The high computing accuracy inherent in digital computing enables error models of the I.N. to be constructed without loss in overall accuracy. This opens up the possibility of mixing at the output of the I.N. in such a manner that three separate navigation outputs are available, Pure I.N., the Nav-Aid and Corrected I.N., and a credibility check made between them. Operating in this manner considerably increases overall system integrity and does away with the distinction between mixed and aided systems.

(b) The flexibility of digital techniques allows the effective gain in the mixing loops to be varied as required by system parameters so that it is operating in the optimum manner at all stages.

(c) Maximum use can be made of the known error characteristics of the I.N. and the Nav-Aids so as to get the best possible correction in a given time.

These advantages can be obtained by using Optimal Filter theory in which corrections are applied at any time on the basis of maximum probability in terms of all the known system characteristics. The Kalman filtering system is one that is particularly suitable for use with digital computers because the information required to make the future best estimate of the I.N. errors is stored in two sets of numbers, the current estimate of the I.N. errors (called the State Vector) and the current estimate of the variances of these errors, and it is not necessary to store the error values occurring each time the I.N. is up-dated.

Figure 4 shows, in block schematic form, the operation of a Kalman Filter network combining I.N. with another Navigation Aid. In the top loop we see the way the current estimate of the I.N. errors (the state vector) is processed by simulated Inertial Platform equations so that they are continuously up-dated in the absence of external information and then added on to the I.N. outputs to provide the current best estimate of the combined system. When Nav-Aid information is available the error between it and the corrected Inertial outputs, after adjustment by a suitable weighting factor, is used to correct the values of these error estimates in the state vector. The weighting factor is computed from the known noise-variance in the Nav-Aid and the variances and covariances of the various elements of the state vector, which are continuously up-dated in the bottom loop from pre-set values.

In this way the weighting factor logic apportions the measured error between the elements of the state vector in the most probable manner on the assumption of normal error distribution.

The pre-set matrix of the variances of these elements are similarly up-dated by I.N. equations with the additional provision for the addition of random changes coming in with the passage of time. The gradual deterioration in the value of the variances is, however, restored each time external information is available, to an extent depending upon the noise variance of the Navigation Aid, thus taking into account the improved certainty with which the errors are known after each correction. This process of continually up-dating the variances used in determining the weighting factor for error correction is equivalent to continuously variable feedback gain in analogue terms, and gives the filter its property of making the best possible use of the information available.

The advantages of the Kalman Filter approach to hybrid I.N. systems can then be summarised as follows:

(a) It will provide optimum correction with any Nav-Aid or combination of Nav-Aids whether operating continuously or intermittently.

(b) The number of I.N. characteristics corrected depends solely on the choice of elements in the state vector. A full air-alignment capability is available with either Doppler or Fixing types of aid.

(c) Besides providing the corrections to the I.N., the computer provides a measure of the credibility of the correction which can be displayed to the operator. This facility, coupled with the ability to preserve the uncorrected I.N. output, makes the system very flexible as a man-machine interface that optimises the use of the individual sensors from the overall integrity aspect.

Another interesting trend in the future will be the increased availability of continuous fixing aids with a world-wide capability, which will probably replace Doppler as the back-up to I.N. Apart from Automatic Astro, VLF navigation aids, such as Omega, and the use of satellite navigation may be mentioned as interesting possibilities for future integration with I.N. which, apart from specialised military applications, may well be competitive on a cost-effectiveness basis with multiplex I.N. installations in the civil field.

5. CONCLUSIONS

Despite the rapid progress which has been made in pure I.N. capability, it has still a long way to go before the complete mission in civil or military terms can be completed with I.N. as the only navigation aid. In terms of en-route navigation, hybrid I.N. systems are likely to have an important role, both for long range civil aircraft and for military maritime reconnaissance. In the terminal area and landing phase, and for weapon aiming purposes, specialised aids will in any case be employed and there are many potential advantages in closely integrating these aids with I.N. The ability of the digital computer, using techniques such as Kalman filtering, to make optimum use of the available sensors at each stage of the mission would appear to make this approach the ideal means of mechanising hybrid I.N. systems in the future.

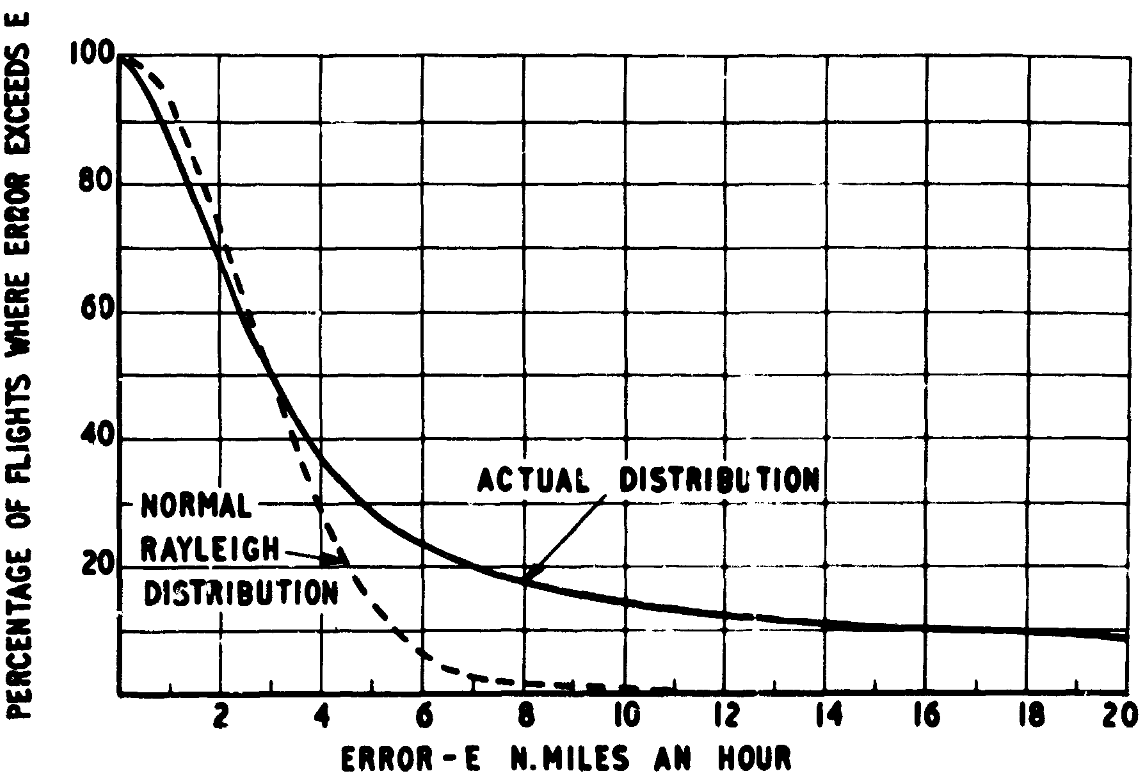


Fig. 1 Error distribution of an Inertial Navigation system

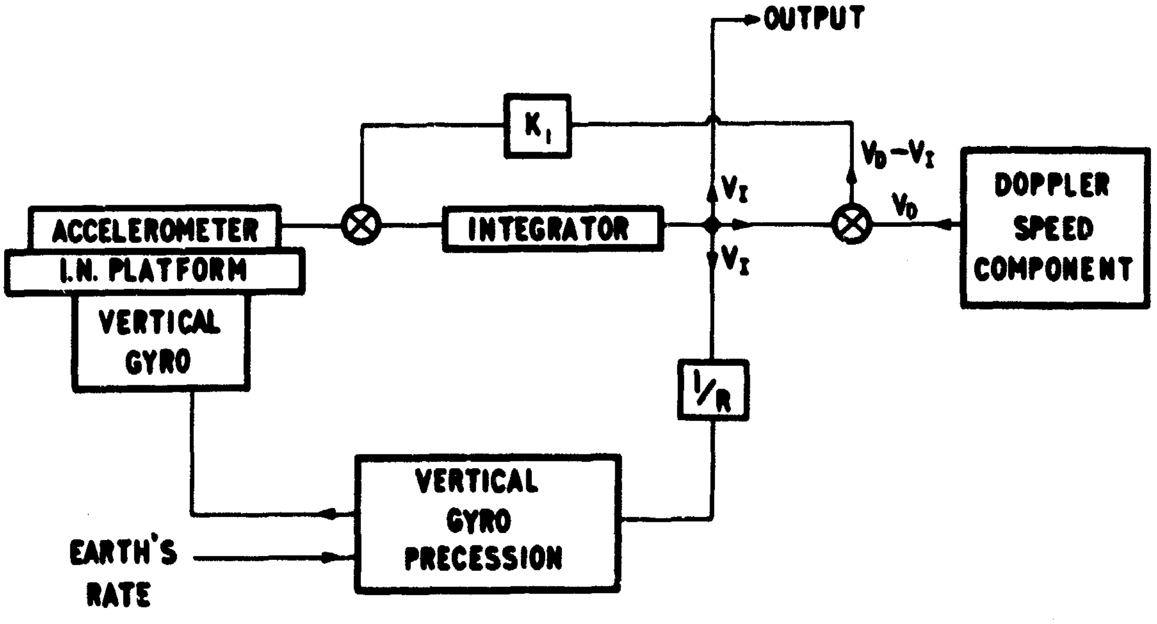
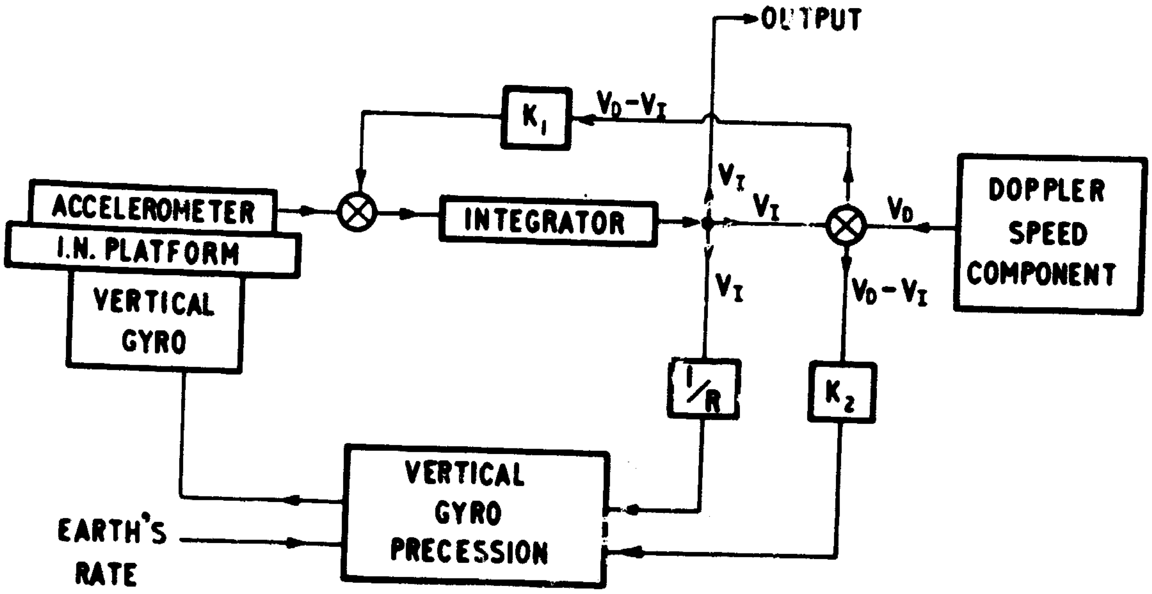


Fig. 2 Doppler Damped Inertia system



$$\overline{V_D - V_I} = \frac{\overline{\Delta D}}{K_2R+1} + \frac{R(\overline{W}+\Omega\cos\lambda\Delta\psi)}{K_2R+1}$$

Fig.3 Doppler Inertia Mixed system

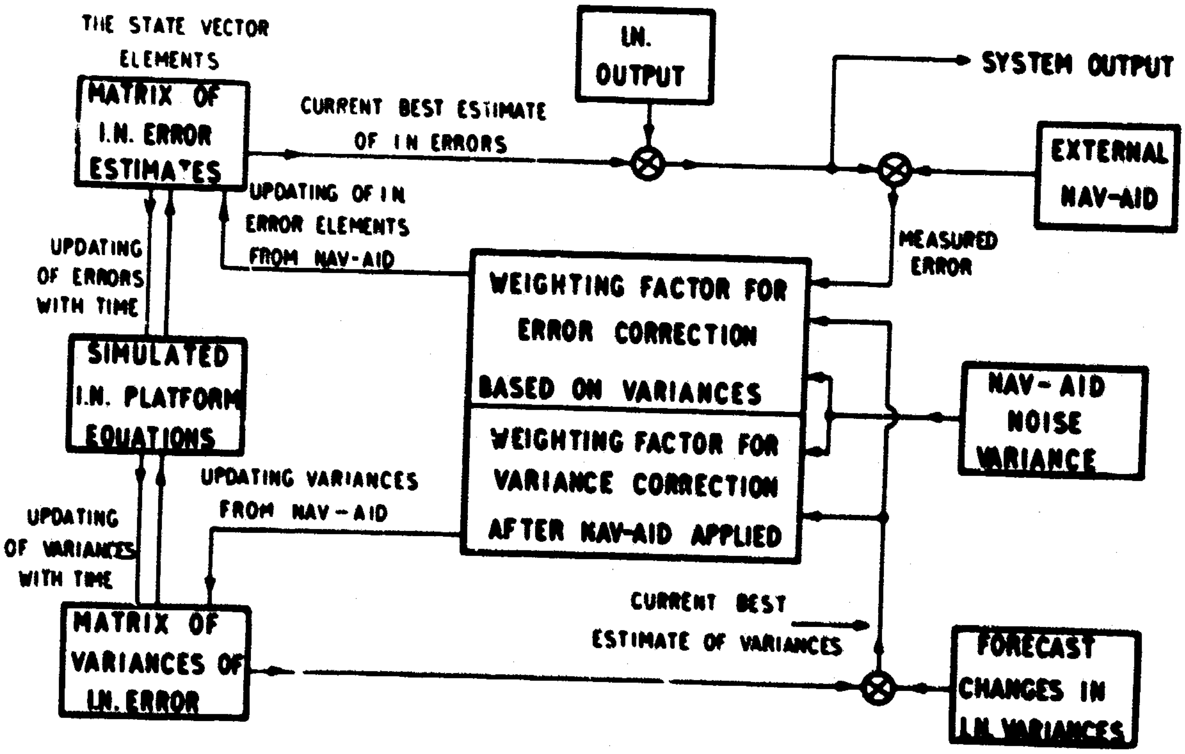

THE STATE VECTOR ELEMENTS
MATRIX OF I.N. ERROR ESTIMATES
I.N. OUTPUT
SYSTEM OUTPUT
CURRENT BEST ESTIMATE OF I N ERRORS
EXTERNAL NAV-AID
UPDATING OF I N ERROR ELEMENTS FROM NAV-AID
MEASURED ERROR
UPDATING OF ERRORS WITH TIME
WEIGHTING FACTOR FOR ERROR CORRECTION BASED ON VARIANCES
SIMULATED I.N. PLATFORM EQUATIONS
NAV-AID NOISE VARIANCE
WEIGHTING FACTOR FOR VARIANCE CORRECTION AFTER NAV-AID APPLIED
UPDATING VARIANCES FROM NAV-AID
UPDATING OF VARIANCES WITH TIME
CURRENT BEST ESTIMATE OF VARIANCES
MATRIX OF VARIANCES OF I.N. ERROR
FORECAST CHANGES IN I.N. VARIANCES

Fig.4 Operation of Kalman Filter network

NAVIGATION PAR INERTIE SUR DE LONGUES DISTANCES

par

B. de Crémiers

Société SAGEM, 6 Avenue d'Iéna, Paris 16^{e}, France

NOTATIONS

(a) *Coordonnées de position*

L, G	latitude et longitude géographiques
L*, G*	latitude et longitude auxiliaires dans le système ayant pour équateur le grand cercle AB dont le pôle est le Nord* (Nord auxiliaire)

(b) *Paramètres angulaires*

ϕ	angle que fait l'équateur auxiliaire avec l'équateur géographique
ψ	angle que fait l'axe de référence OX de la plateforme avec la direction du Nord géographique
α	angle que fait l'axe de référence OX de la plateforme avec la direction de l'Est auxiliaire (Remarque: l'angle α est une constante quelle que soit la route suivie pour aller de A à B).

(c) *Vitesses et accélérations linéaires*

V_x, V_y	composantes de la vitesse sol horizontale sur les axes OX et OY
V_{Ix}, V_{Iy}	intégrales des mesures accélérométriques
C_x, C_y	accélérations de Coriolis
V_N, V_E	composantes de la vitesse sol horizontale sur les axes géographiques
V_z	vitesse verticale, comptée positivement vers le bas.

(d) *Vitesses angulaires et taux de précession*

L'^*, G'^*	vitesses angulaires en latitude et longitude auxiliaires
ω	vitesse de rotation de la terre
ω_x, ω_y	taux de bouclage de Schuler
ω_z	taux de convergence des méridiens auxiliaires
D_{0x}, D_{0y}, D_{0z}	compensations de la dérive des gyros
$\Omega_x, \Omega_y, \Omega_z$	taux de précession commandés

(e) *Autres symboles*

S	angle que fait la route suivie avec le Nord géographique
R_p, R_q	rayons de courbure principaux (rapportés aux axes géographiques)
$R_{p0}, R_{q0}, \epsilon_p, \epsilon_q$	constantes de l'ellipsoïde terrestre:

$$
\begin{aligned}
R_{p0} &= 6\,357\,580 \text{ m} \\
R_{q0} &= 6\,388\,900 \text{ m} \\
\epsilon_p &= 0{,}0050288 \\
\epsilon_q &= 0{,}0016762
\end{aligned}
$$

R	rayon de courbure normale
T	inverse de la torsion géodésique
h	altitude au dessus de l'ellipsoïde de référence
ρ, τ	rayon de courbure normale et inverse de la torsion géodésique rapportés à l'altitude h .

NAVIGATION PAR INERTIE SUR DE LONGUES DISTANCES

B. de Crémiers

1. INTRODUCTION

On sait que l'avion de transport supersonique "Concorde" sera équipé de systèmes de navigation par inertie. Les systèmes destinés à équiper les avions prototypes sont actuellement construits par la Société française SAGEM et la Société britannique Ferranti. C'est à cette occasion que la Société SAGEM a eu la charge d'étudier l'application des systèmes à inertie à la navigation des avions sur de longues distances.

Il était nécessaire de rechercher les meilleures solutions compatibles avec certaines caractéristiques de l'utilisation, et notamment les suivantes:

- autonomie de la recherche du Nord avant l'envol, en évitant si possible le recours à des compensations ou calibrations préalables
- navigation universelle, c'est-à-dire pouvant couvrir toutes les régions du globe, et en particulier les zones polaires
- découpage des routes suivies en tronçons en nombre quelconque, et possibilité de diversion en dehors de ces tronçons
- économie souhaitable des capacités du calculateur de bord.

La nécessité de la couverture des zones polaires a donné à l'étude son point de départ: en effet, un fonctionnement correct de la plateforme inertielle lorsque l'avion navigue au voisinage du pôle exige que cette plateforme ne soit pas "asservie au Nord"; cela est dû au fait que le taux de précession que l'on peut matériellement lui imposer autour de l'axe d'azimut est limité.

Dès lors, la question s'est posée du choix de l'orientation initiale lors de l'alignement de la plateforme avant l'envol: la recherche du Nord peut être effectuée en effet soit par la méthode classique du gyrocompas soit par une méthode de calcul. C'est cette dernière méthode qui a été choisie et développée afin d'en tirer, comme on le verra plus loin, les deux avantages principaux suivants:

(a) Le mode d'alignement est unique.

(b) La correction de la dérive des gyros est entretenue sans qu'il soit nécessaire de procéder à des calibrations périodiques.

En ce qui concerne la navigation, on a choisi d'opérer tous les calculs dans un système de coordonnées sphériques. Il est apparu en effet que les exigences de l'utilisation du système de navigation nécessitaient la mise en place d'un moyen de résolution de triangles sphériques.

Cette résolution est opérée par un sous-programme du calculateur qui constitue en même temps une matrice de transformation de coordonnées sphériques quelconques. On a donc considéré que le passage d'un système de coordonnées sphériques à un autre est une opération banale effectuée sur appel du sous-programme, et que le nombre de systèmes de coordonnées sphériques auxiliaires utilisés n'est pas limité. Dès lors, les calculs de navigation peuvent, en toutes circonstances, y compris dans les zones polaires, être effectués dans un système de coordonnées shpériques auxiliaires convenablement choisi, et cela toujours de la même façon. Ainsi n'y a-t-il qu'un seul mode de calculs de navigation, et la mémoire d'instructions du calculateur s'en trouve allégée.

En résumé, le système de navigation à inertie qui a été étudié en vue de son utilisation sur de longues distances est caractérisé par le fait remarquable qu'il ne comporte, en ce qui concerne les calculs, que deux modes d'opération: un mode alignement et un mode navigation.

La description de ces modes, ainsi que leurs avantages et leurs particularités font l'objet des paragraphes qui suivent.

2. MODE ALIGNEMENT

2.1 Description

La plateforme, ou plus exactement le coeur de la plateforme, n'est pas asservi au Nord; en particulier, lors de la mise en route, son orientation en azimut est quelconque, et est maintenue fixe à sa valeur initiale pendant l'alignement de la plateforme.

On admet que le calculateur connaît la latitude du lieu, et a conservé en mémoire les valeurs des dérives propres des gyros mesurées antérieurement (d_{x0}, d_{y0}, d_{z0}) .

A la mise en route de l'equipement, un calage préliminaire sur synchros est opéré en horizontalité, en prenant comme référence roulis et tangage nuls; d'autre part le synchro de cap de la plateforme est recopié par un répétiteur.

Après chauffage convenable et mise en route des gyros, le répétiteur de cap plateforme est immobilisé, le calculateur impose aux gyros les taux de précession suivants;

- d_{x0} sur le gyro de verticale Gx dont l'axe d'entrée fait, avec le méridien, un angle ψ .
- d_{y0} sur le gyro de verticale Gy dont l'axe d'entrée fait, avec celui du gyro Gx , un angle de $\pi/2$.
- Ω sin L - d_{z0} sur le gyro d'azimut Gz .

D'autre part, les bouclages suivants sont opérés:

(a) *En horizontalité*

L'accéléromètre Ax est bouclé sur le gyro Gy , et l'accéléromètre Ay sur le gyro Gx , selon le diagramme de principe de la Figure 1. Les paramètres de ces boucles sont choisis pour leur donner une période naturelle de l'ordre de 6 minutes et un amortissement de l'ordre de 0,7.

Il est facile de montrer qu'à la position d'équilibre, la sortie de l'accéléromètre Ax est nulle et la sortie de l'intégrateur mesure la vitesse angulaire sentie par le gyro Gy , c'est-à-dire la somme de la composante de la rotation terrestre sur son axe d'entrée et d'une dérive propre résiduelle d_y (différence entre la dérive propre actuelle et la dérive compensée antérieurement d_{y0} , c'est-à-dire "Dérive de jour à jour").

La position d'équilibre des boucles d'horizontalité est définie par les relations suivantes:

$$\dot{\varepsilon}_x = 0$$

$$-Uy = -\Omega \cos L \sin \psi + d_y$$

$$\dot{\varepsilon}_y = 0$$

$$-Ux = \Omega \cos L \cos \psi + d_z \,.$$

D'autre part le calculateur effectue les opérations:

$$\psi_1 = -\mathrm{arc\ tg}\ (Uy/Ux)$$

$$d_1 = \sqrt{(Ux^2 + Uy^2)} - \Omega \cos L$$

$$d_{x1} = + d_1 \cos \psi_1$$

$$d_{y1} = - d_1 \sin \psi_1 \,.$$

Ces résultats ont la signification suivante: ψ_1 est la valeur de l'angle que fait avec le méridien la direction OX de la plateforme, à une erreur près ϵ, dont le calcul est indiqué dans l'annexe, et dont la valeur prévisible est pratiquement égale au rapport de la dérive de jour à jour des gyros de verticale à la composante horizontale de la rotation terrestre. d_1 est la valeur de la dérive de jour à jour des gyros de verticale, en projection sur l'axe ON, et constitue donc un test de confiance dans les gyros.

d_{x1} et d_{y1} sont les termes complémentaires de compensation qu'il faut ajouter respectivement à d_{x0} et d_{y0} pour annuler la composante Nord de la dérive des deux gyros de verticale.

(b) *En azimut*

La liaison synchro de cap plateforme sur le répétiteur immobilisé est bouclée sur le gyro d'azimut selon le diagramme de principe de la Figure 2. Les paramètres de cette boucle sont choisis pour lui donner une période naturelle de l'ordre de 6 minutes et un amortissement de l'ordre de 0,7.

A la position d'équilibre, la sortie de l'intégrateur mesure la dérive de jour à jour d_{z1} du gyro Gz.

2.2 Mode opératoire

En mode opératoire normal, le système est placé en "Alignement" après une séquence automatique, réalisant le chauffage de la plateforme, le calage préliminaire sur synchros, et la mise en route des gyros.

Après 6 minutes environ de fonctionnement sur le mode alignement, le calculateur présente des résultats sur des affichages appropriés:

- le cap de l'avion C (obtenu par différence du "Cap plateforme" Cp, délivré par son synchro, et de l'angle calculé ψ_1: $C = Cp - \psi_1$),

- les dérives de jour à jour des gyros d_1 et d_{z1}.

Le pilote est alors en mesure de porter une appréciation sur le fonctionnement du système. En effet, le cap C peut être immédiatement comparé au cap magnétique, en tenant compte de la déclinaison locale; d'autre part, les dérives affichées d_1 et d_{z1} doivent être inférieures aux tolérances admises pour les dérives de jour à jour.

Lors du passage en mode "Navigation" les valeurs des termes d_{x1}, d_{y1} et d_{z1} seront automatiquement prises en compte par le calculateur pour la mise à jour des termes d_{x0}, d_{y0}, d_{z0} conservés en mémoire.

Si le pilote dispose d'un temps suffisant, il pourra, au lieu de passer en mode "Navigation", "répéter" le mode alignement. L'ordre de répétition déclenche les opérations suivantes:

- les termes d_{x1}, d_{y1} et d_{z1} sont pris en compte par le calculateur,
- la plateforme est décalée de 90° en azimut (en imposant au gyro d'azimut une précession rapide),
- un nouvel alignement est opéré.

Le calculateur élabore les nouveaux résultats:

$$\psi_2 = -\text{arc tg } (Uy/Ux)$$

$$d_2 = \sqrt{(Ux^2 + Uy^2)} - \Omega \cos L$$

$$d_{x2} = + d_2 \cos \psi_2$$

$$d_{y2} = - d_2 \sin \psi_2$$

Ces nouveaux résultats ont la signification suivante: ψ_2 est la valeur de l'angle que fait avec le méridien la direction OX de la plateforme; mais il n'y a plus d'erreur ϵ, la dérive d_1 ayant été corrigée d_2 est la valeur de la dérive de jour à jour des gyros de verticale, en projection sur l'axe ON, et, du fait de la rotation de 90°, on a

$$\vec{d_1} + \vec{d_2} = \vec{d_x} + \vec{d_y} .$$

d_{x2} et d_{y2} sont les termes complémentaires qu'il faut ajouter respectivement à d_{x1} et d_{y1} pour annuler totalement la dérive des deux gyros de verticale.

Ce mode peut, bien entendu, être répété à nouveau, la plateforme étant placée en alignement après des décalages successifs de 90°. L'observation du cap calculé et des mesures successives des dérives d et d_z qui, à partir du troisième alignement doivent être nulles, ou inférieures aux tolérances admises pour les dérives à court terme, constitue un test de bon fonctionnement au sol, le délai nécessaire pour passer en mode navigation étant toujours au plus égal à la durée minimum d'un alignement (10 minutes environ).

2.3 Avantages

La méthode d'alignement précédemment décrite possède un certain nombre d'avantages remarquables:

(a) Il n'y a qu'un seul mode d'alignement, il en résulte une simplicité des commutations et des programmes de calcul.

(b) L'orientation de la plateforme en azimut étant quelconque, la connaissance approximative de la direction du Nord n'est pas nécessaire pour favoriser les conditions initiales de l'alignement.

(c) Le mode alignement peut être répété à volonte - avec décalage de la plateforme de 90° en azimut - pour parfaire la correction des dérives des gyros, sans qu'il soit nécessaire de connaître *a priori* le délai dont on dispose; le délai nécessaire pour passer en mode navigation est toujours au plus égal à la durée minimum d'un alignement.

(d) La répétition du mode alignement constitue le meilleur test de fonctionnement au sol.

(e) En mode opératoire normal, comportant un seul mode alignement, la correction de la dérive des gyros est entretenue par le fait du caractère aléatoire de l'orientation de la plateforme en azimut, lors des mises en route successives. Il en résulte qu'il n'y a pas lieu de faire une révision périodique de la compensation des gyros, celle-ci étant entretenue par l'utilisation opérationnelle du système.

3. MODE NAVIGATION

3.1 Description

3.1.1

La méthode utilisée pour le mode navigation repose sur deux idées de base qui ont été inspirées par les conditions normales d'utilisation:

(a) La première idée de base est que l'avion a l'intention d'aller d'un point origine A à un point de destination B, en suivant un arc de grand cercle, et qu'il importe autant de connaître à tout instant la position par rapport à ces points - c'est-à-dire l'écart latéral par rapport à la route à suivre, et la distance du point de destination - que la position exprimée en latitude et longitude géographiques. Cette notion montre que le système est orienté vers la fonction "guidage" de l'avion, plutôt que vers la seule fonction de "navigation", cette dernière pouvant se réduire à la connaissance de la position présente, de la route actuelle et de la vitesse sol.

Il faut ajouter que les points A et B sont définis par leur latitude et longitude géographiques, qu'ils sont en nombre quelconque, le plan de vol pouvant comporter un grand nombre de tronçons successifs, et que leurs coordonnées sont soit conservées en mémoire dans le calculateur, soit introduites manuellement à volonté; enfin, la position présente peut en outre être choisie comme point origine pour une destination quelconque.

(b) La deuxième idée de base est que l'arc de grand cercle AB étant choisi comme équateur d'un système de coordonnées sphériques auxiliaires, les calculs de navigation seront toujours effectués en totalité dans ce système, y compris les intégrations nécessaires à la connaissance de la position, la latitude et la longitude géographiques étant obtenues finalement par transformation de coordonnées. Cependant, on calcule les composantes Nord et Est de la vitesse sol afin d'en déduire la route suivie, élément utilisé à titre d'information, et non comme paramètre de guidage.

3.1.2

Pour mieux décrire les équations du mode navigation qui sont traitées par le calculateur, il faut tout d'abord préciser ce que l'on attend des sorties du calculateur:

(a) *Conduite de la plateforme*
La calculateur devra fournir les taux de précession à imposer aux gyroscopes de la plateforme afin d'une part de maintenir celle-ci horizontale et d'autre part de la faire précessionner en azimut à un taux égal à celui de la rotation du méridien auxiliaire (grand cercle joignant la position présente au pôle de l'arc AB).

(b) *Conduite de l'avion*
Le calculateur devra fournir les coordonnées de position de l'avion par rapport à l'arc AB à parcourir, ainsi que la transformation de ces coordonnées en latitude et longitude.

Le calculateur devra fournir en outre divers facteurs, fonctions de la position présente, nécessaires à l'élaboration des taux de précession de la plateforme, ainsi que la route suivie.

3.1.3

Les équations traitées par le calculateur sont explicitées dans les Figures 3 et 4.

On remarquera que la charge de travail du calculateur a été répartie en deux groupes:

(a) un groupe de calculs est opéré à fréquence rapide, soit 10 Hz; ces calculs sont tous des calculs de vitesses : vitesse horizontale, taux de bouclage des boucles de Schüler, vitesses angulaires en latitude et longitude auxiliaires, convergence des méridiens auxiliaires, taux de précession à imposer aux gyros de la plateforme.

(b) l'autre groupe de calculs est opéré à fréquence lente; une fréquence de 0,25 Hz serait suffisante, mais la capacité du calculateur autorise en fait une fréquence de 2 Hz. Ces calculs sont surtout des calculs de position présente et de facteurs fonctions de la position présente (fonctions sinus et cosinus de la latitude auxiliaire, de la latitude géographique, de l'angle φ de référence de la plateforme par rapport au Nord, rayon de courbure et torsion de la trajectoire engendrée par le vecteur vitesse horizontale). On y ajoute le calcul des accélérations de Coriolis, ainsi que celui des composantes Nord et Est de la vitesse afin seulement d'en déduire la route suivie.

3.2 Remarques

3.2.1 Définition de la navigation sur un arc de grand cercle

Comme on l'a dit plus haut, les calculs de navigation sont effectués dans un système de coordonnées sphériques dont l'équateur est le grand cercle de la route à suivre. Il est intéressant, à ce sujet, de faire une remarque sur la signification des équations utilisées.

La loi de guidage de l'avion sur le grand cercle à suivre peut s'écrire très simplement sous la forme

$$L'^{*} = 0 ,$$

soit $\dfrac{\omega_x}{\omega_y} = -\mathrm{tg}\alpha$, α étant une constante caractérisant le grand cercle à suivre.

Si on se reporte aux équations du bouclage de Schüler qui font apparaître que ω_x/ω_y est différent de $-V_y/V_x$ en raison de la présence du terme de torsion géodésique, on en déduit que l'arc dont la tangente est $-V_y/V_x$ est non pas une constante, mais une variable, c'est-à-dire que le "Cap plateforme" n'est pas constant.

Il s'ensuit que la loi de guidage adoptée,

$$L'^{*} = 0 ,$$

est légèrement différente de la loi, familière aux navigateurs, selon laquelle la route orthodromique est suivie en maintenant à une valeur constante le cap fourni par un système conventionnel de référence gyroscopique directionnelle (corrigé de la rotation terrestre et non de la convergence des méridiens).

On a calculé que la différence entre ces deux lois se traduit par un écart latéral de 6 nautiques environ sur le parcours Paris - New York, et un écart négligeable le long de la route.

3.2.2 *Navigation sans Destination*

On a dit qu'une des idées de base qui ont servi à l'établissement des équations de la navigation est l'intention d'aller d'un point A à un point B. Or, il est très concevable également de prendre l'air sans qu'une destination soit à priori imposée.

En ce cas, on choisira comme destination arbitraire un point quelconque du méridien origine, le pôle Nord pour plus de simplicité, ce qui revient à adopter un système de coordonnées sphériques auxiliaires dont le méridien origine est l'équateur.

Compte tenu du fait que la convergence des méridiens de ce système auxiliaire est toujours entretenue, il sera possible de naviguer sans destination dans toute la zone comprise entre les parallèles auxiliaires $\pm 60^{\circ}$ c'est-à-dire au moins jusqu'à 3600 nautiques du point de départ. Avant que cette distance soit atteinte, il est plus que probable qu'une destination aura pu être choisie et qu'ainsi on aura pu adopter un autre système auxiliaire ayant pour équateur le grand cercle joignant la position présente à l'instant du choix, à la destination choisie.

3.3 Avantages

La méthode utilisée pour le mode navigation présente les avantages suivants:

(a) Le mode navigation est universel. Les calculs sont en effet toujours les mêmes, quelles que soient les destinations ou les régions survolées, y compris les zones polaires.

(b) Du fait de l'unicité du mode navigation, le programme d'instructions du calculateur est allégé; il l'est d'autant plus que le sous-programme de transformation de coordonnées sphériques est en outre utilisé à d'autres fins.

4. CONCLUSION

On a montré brièvement pourquoi et comment un système de navigation à inertie a été conçu en vue de son utilisation sur de longues distances.

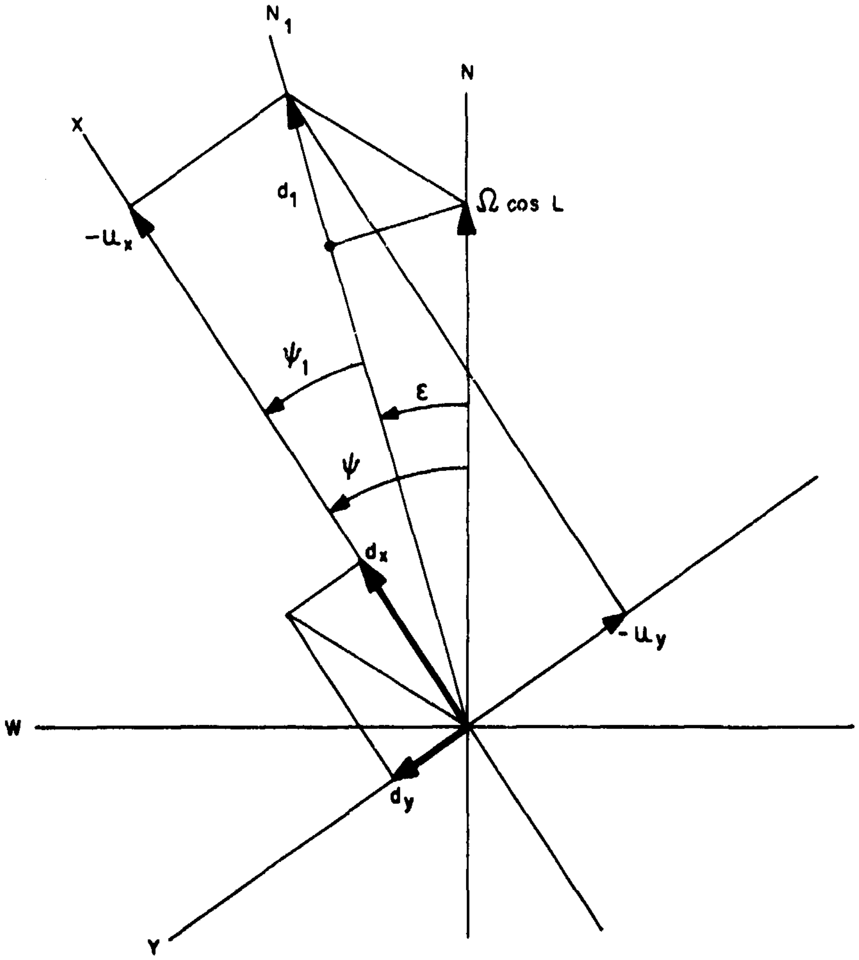
Les avantages du système décrit sont manifestes; les deux seuls modes d'opération sont

- le mode alignement, qui permet, outre sa fonction principale, d'entretenir la correction de la dérive des gyroscopes, rendant ainsi inutiles les calibrations périodiques,
- le mode navigation, qui non seulement permet de couvrir de façon universelle toutes les régions du globe, mais aussi est particulièrement bien adapté à la fonction "guidage" de l'avion.

Il faut ajouter qu'un tel système n'a pu être conçu qu'en faisant appel à l'aide puissante d'un calculateur numérique dont le rôle est évidemment primordial, tant dans la recherche du Nord, lors de l'alignement, que dans l'élaboration des signaux de guidage lors de la navigation.

ANNEXE

CALCULS RELATIFS AUX DERIVES DES GYROS Gx ET Gy



La dérive d_x du gyro Gx est équivalente à un vecteur rotation d_x porté par OX ; de même la dérive du gyro Gy peut être représentée par le vecteur d_y porté par OY .

La résultante de d_x et d_y , ajoutée au vecteur $\Omega \cos L$ porté par ON donne la direction du Nord apparent ON_1 avec lequel l'axe OX fait l'angle ψ_1 défini par

$$\psi_1 = -\text{arc tg } (Uy/Ux)$$

avec

$$-Uy = -\Omega \cos L \sin \psi + d_y$$

$$-Ux = \Omega \cos L \cos \psi + d_x \; .$$

En projection sur l'axe OW_1 perpendiculaire à ON_1 , la résultante du vectuer $\vec{d_x} + \vec{d_y}$ a pour valeur

$$d_x \sin \psi_1 + d_y \cos \psi_1 \; .$$

L'angle ϵ , angle que fait le Nord apparent avec le Nord vrai, est un petit angle, et on peut écrire

$$\epsilon = \frac{d_x \sin \psi_1 + d_y \cos \psi_1}{\Omega \cos L}$$

En projection sur l'axe ON_1, la résultante du vecteur $\vec{d_x} + \vec{d_y}$ a pour valeur

$$d_1 = d_x \cos \psi_1 - d_y \sin \psi_1$$

dont une autre expression est, pour ϵ petit,

$$d_1 = \sqrt{(Ux^2 + Uy^2)} - \Omega \cos L .$$

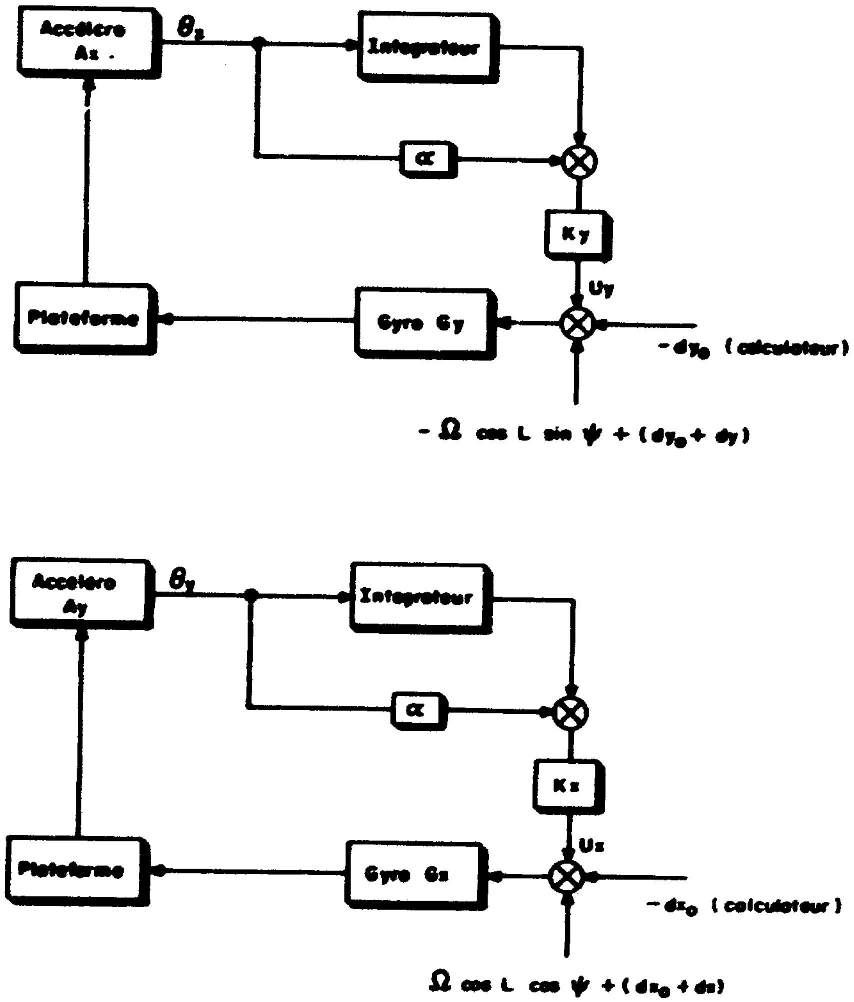


Fig. 1 Boucles d'alignement en horizontalité

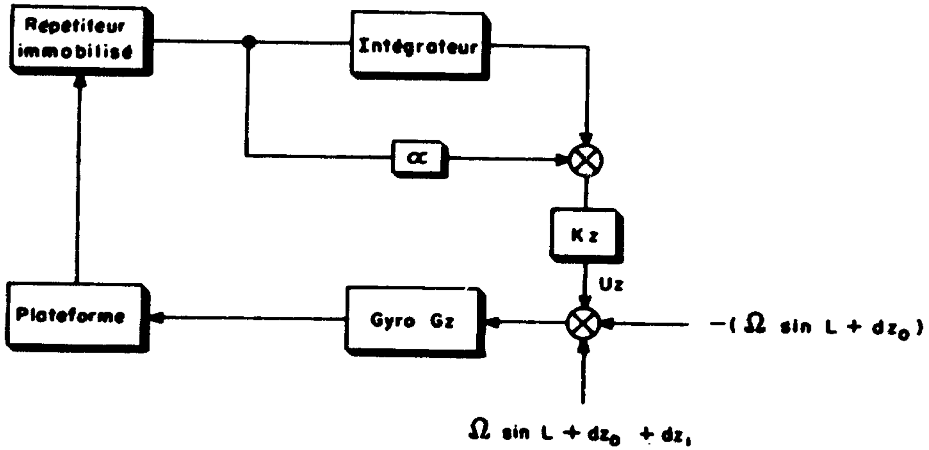


Fig. 2 Boucle de mesure de la dérive du gyro Gz

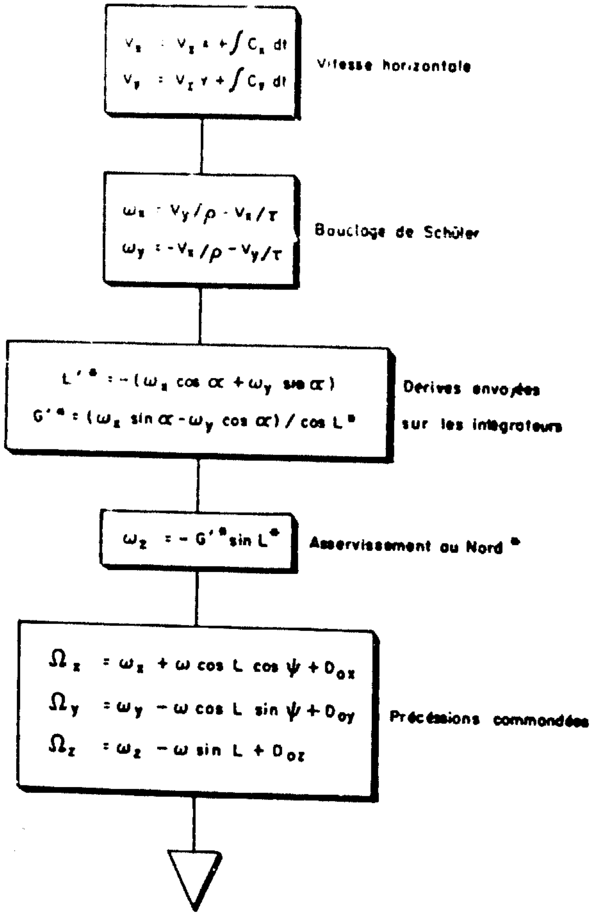

$V_x = V_{x}x + \int C_x\,dt$
$V_y = V_{I}y + \int C_y\,dt$
Vitesse horizontale
$\omega_x = V_y/\rho - V_x/\tau$
$\omega_y = -V_x/\rho - V_y/\tau$
Bouclage de Schüler
$L'^{*} = -(\omega_x \cos\alpha + \omega_y \sin\alpha)$
$G'^{*} = (\omega_x \sin\alpha - \omega_y \cos\alpha)/\cos L^{*}$
Derives envoyées sur les intégrateurs
$\omega_z = -G'^{*} \sin L^{*}$
Asservissement au Nord *
$\Omega_x = \omega_x + \omega\cos L\cos\psi + D_{ox}$
$\Omega_y = \omega_y - \omega\cos L\sin\psi + D_{oy}$
$\Omega_z = \omega_z - \omega\sin L + D_{oz}$
Précessions commandées

Fig.3 Groupe de calculs à fréquence rapide

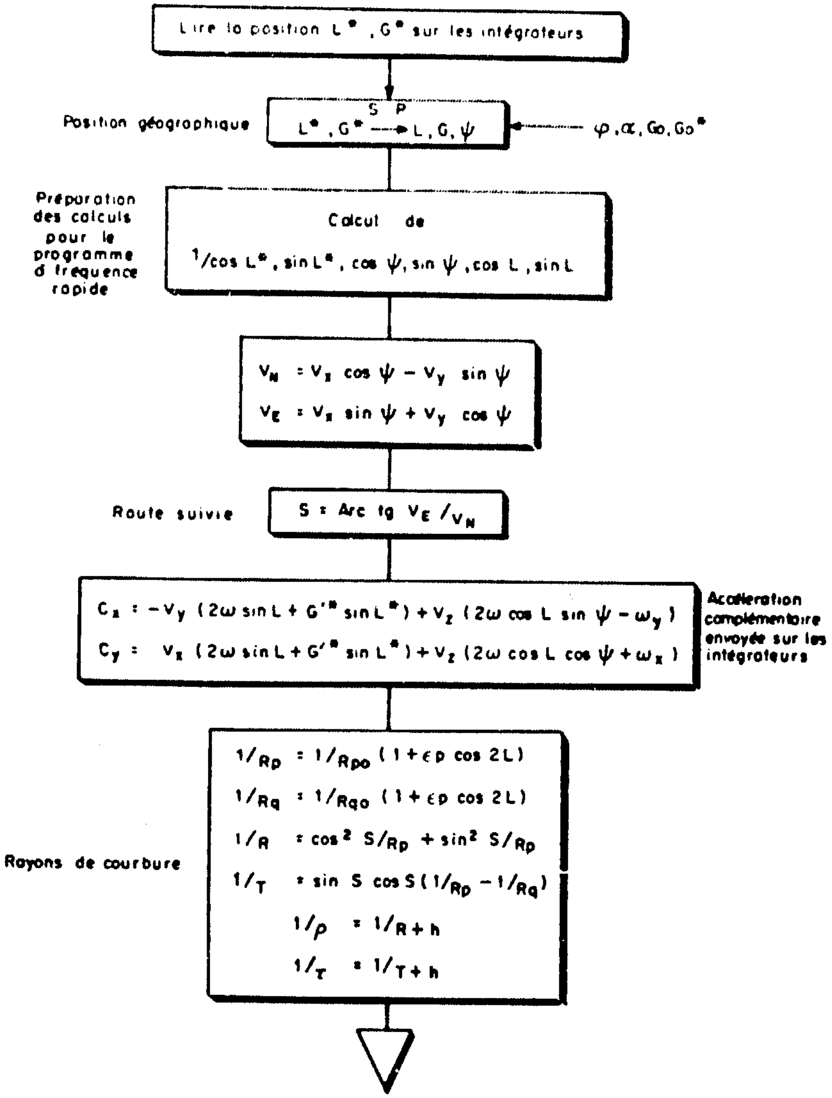

Lire la position L*, G* sur les intégrateurs
Position géographique
$L^{*}, G^{*} \xrightarrow{S\ P} L, G, \psi$
$\varphi, \alpha, Go, Go^{*}$
Préparation des calculs pour le programme d fréquence rapide
Calcul de
$1/\cos L^{*}, \sin L^{*}, \cos\psi, \sin\psi, \cos L, \sin L$
$V_N = V_x \cos\psi - V_y \sin\psi$
$V_E = V_x \sin\psi + V_y \cos\psi$
Route suivie
$S = \mathrm{Arc\ tg}\ V_E/V_N$
$C_x = -V_y(2\omega\sin L + G'^{*}\sin L^{*}) + V_z(2\omega\cos L\sin\psi - \omega_y)$
$C_y = V_x(2\omega\sin L + G'^{*}\sin L^{*}) + V_z(2\omega\cos L\cos\psi + \omega_x)$
Accélération complémentaire envoyée sur les intégrateurs
Rayons de courbure
$1/R_p = 1/R_{po}(1 + \epsilon p \cos 2L)$
$1/R_q = 1/R_{qo}(1 + \epsilon p \cos 2L)$
$1/R = \cos^2 S/R_p + \sin^2 S/R_p$
$1/T = \sin S \cos S(1/R_p - 1/R_q)$
$1/\rho = 1/R+h$
$1/\tau = 1/T+h$

Fig.4 Groupe de calculs à fréquence lente

Best Available Copy

COST-OF-OWNERSHIP PHILOSOPHY APPLIED TO INERTIAL NAVIGATION SYSTEMS

by

W.J.Laubendorfer
R.V.Plank
E.J.DeNezza

Air Force Avionics Laboratory
Holloman Air Force Base, New Mexico, USA

SUMMARY

The advent of inertial navigation systems is approaching its twentieth anniversary and during this period major emphasis has been placed on performance. This has required the design and manufacture of equipment in which performance has been the dependent variable to be maximized. In many instances, state of the art designs were needed on almost a continuous basis. It is the contention of this paper that, while performance should be given prime emphasis, the economics of using inertial navigation equipment should also be given serious consideration. The Air Force previously has fostered the joint cost-performance approach through the establishment of Detachment 1 of the AF Avionics Laboratory at Holloman AFB, New Mexico, along with other stimuli to the various potential contractors for inertial navigation systems. Some members of the industrial inertial community now are considering the use of their equipment from not only a performance but from also a cost totality aspect. For many years, over-simplified expressions were applied to determine system acceptability - acceptable or non-acceptable - with little regard to system reliability. Consequently, what can easily happen, and what has happened in the past, is that equipments which appear very attractive from an initial cost standpoint, can require significant funding in support of operational use. This paper supports the need for a total performance (including reliability and maintainability) cost-of-ownership approach toward inertial navigation system design and selection. In addition, a concept entitled COPE - an acronym for Cost-of-Ownership Performance is presented, defined and applied to inertial navigation system design and selection.

COST-OF-OWNERSHIP PHILOSOPHY
APPLIED TO INERTIAL NAVIGATION SYSTEMS

W.J.Laubendorfer, R.V.Plank and E.J.DeNezza

1. INTRODUCTION

How should a user procure inertial equipment to provide the best combination of economy and performance? Without question the application of two disciplines is necessary, the all too familiar state-of-the-art with regard to performance and the less familiar state-of-the-economy with respect to cost. For many years these two disciplines were applied by first determining technical acceptability and then the low bidder with respect to acquisition cost. It is our contention that this procurement philosophy is an oversimplified application of these two disciplines. All too often this procurement philosophy, when applied, has resulted in the Air Force buying equipment which appeared very attractive from an initial cost standpoint but required excessive funding for operational support.

We hope to show that the Technical Acceptability - Low Bidder philosophy should be replaced by a more comprehensive Performance - Cost-of-Ownership philosophy. In essence, we maintain that

(i) The state-of-the-art with regard to performance must include not only accuracies, reaction times, power, weight, size, etc, but reliability and maintainability.

(ii) The state-of-the-economy with respect to cost must include not only initial costs but complete user cost or cost-of-ownership.

We hope to develop and combine these disciplines into a quantitative, usable tool that, if applied at program start, will not only provide a means of selecting the best system, but will be an instrument for efficient and effective design.

2. TECHNICAL ACCEPTABILITY - LOW BIDDER PHILOSOPHY

All too often, once a system is considered technically acceptable, it generally is procured from the lowest bidder.

To illustrate what may happen, let us consider the following hypothetical situation:

Inertial Systems A and B are considered technically acceptable and characterized by Table I

TABLE I

	System A	*System B*
Acquisition cost ($)	72,500	100,000
Inertial Measuring Unit (IMU) cost ($)	35,000	48,000
IMU MTBF (hours)	120	400
(This does not include any prorated costs for tooling, training, spares, administrative support, inventory, etc.).		

Under the aforementioned procurement philosophy, the user would probably procure inertial System A. On the surface, this appears to be a reasonable choice. However, let us analyze this decision with a very simple cost-of-ownership model using the IMU.

Assume:

(i) An aircraft utilization rate of 20 hours per week.

(ii) A depot repair cost of $8,000 per IMU (System A) and $9,000 per IMU (System B).

(iii) A depot repair cycle for both IMUs of ten weeks.

(iv) A base repair cycle (pipeline time*) for both IMUs of two weeks.

(v) A depot/base repair ratio of 2:1 (i.e. for every three IMU failures, two are repaired at the depot and one on the base).

Table II summarizes the results of Figure 1.

TABLE II

	System A	*System B*
No. of IMU's required for 1 aircraft	3	2
No. of depot repairs (1 yr)	5	2
(3 yr)	15	5

Computing cost of ownership in this model as

C.O. = (Cost per IMU × No. IMUs required) + (No. of Depot Repairs × Cost of Depot Repairs) ,

the cost-of-ownership versus years of service is shown in Table III.

TABLE III

Years/IMU	*A*	*B*
1	$145,000	$114,000
3	$225,000	$141,000

Thus, when translated into the meaningful terms of user costs, the selection of System A indeed becomes the poorer economic selection. Further analysis of the IMU's utilization, as presented in Figure 2, shows the number of systems required increases with aircraft use rate. It is of interest to note that if the depot/base repair ratio is changed incrementally from 1:1 and then finally all depot, the net result is that the same number of IMU's is required; however, the margin of availability is significantly decreased. The margin of availability is defined as the number of system weeks a spare IMU is available prior to an in-use IMU failure. For the all-depot case, the margin of availability is approximately two system weeks. For the 2:1 case, the margin varies from 8.5 system weeks to two system weeks. The user must determine the acceptability of this margin.

---

* Pipeline time is defined as the elapsed time to repair defective equipment, including transportation time.

The above analysis, while hypothetical, is based upon realistic data on repair cycles, repair costs, depot/base repair ratios, etc. and we feel it clearly illustrates the limitations of the technical acceptability-low bidder procurement philosophy.

We propose that technical acceptability be replaced by total performance (i.e. include MTBF, maintainability, etc.) and that serious consideration be given to the elimination of the low bidder philosophy in favor of a low cost-of-ownership philosophy. We advocate this philosophy not only for selection but for the design of future inertial navigation equipment. This approach would result in the identification, collection and analysis of all costs from equipment design inception to obsolescence and phase out. The need for selecting equipment on this basis is acute.

3. PERFORMANCE - COST-OF-OWNERSHIP PHILOSOPHY

It is now necessary to present in more specific terms what we feel should be considered as part of performance and cost-of-ownership.

3.1 Performance

There are many specifications which can be applied to the procurement of equipment. Probably these specifications increase in direct proportion to equipment complexity and environmental use. Should it be required to procure a sphere of 7075 T6 aluminum, measuring three inches in diameter with a sphericity of 50 u" at 72°F having a 16 u" r.m.s. finish, the possible specifications are few. The requirements have been specified fully and can be demonstrated adequately by available measuring equipment. But notice that no requirements have been stipulated for any aspect of time or for use in any environment. A procurement as simple as this one appears would become significantly complex should a time and environmental aspect be included. When a user desires to procure equipment as complex and expensive as an inertial navigation system, the specifications are many and equally complex. The specifications are either the complete responsibility of the supplier nor the user. However, there must be a beginning. It must start with the user, who must be in a position to describe in intelligible terms his performance needs and economic requirements. It is here that realistic, achievable MBTF and maintainability requirements must be specified. His performance needs should not be directed toward the best obtainable, but rather toward that which is sufficient for the task at hand when a reasonable growth factor is applied. This requires personal judgment which must be based on experience and judiciously applied. Furthermore, the method of demonstrating performance characteristics must be included with performance models and mechanizations. Certainly this places a significant technical burden on the user, but that is exactly where it belongs.

3.2 Cost-of-Ownership

The assignment of cost items is more complex and one in which significant latitude can be expected unless meticulously defined and continuously controlled. This area is probably more of a joint effort between user and supplier. While specific cost breakdowns may vary, we believe you will agree that all costs can be included under four general categories: materials, people, facilities and transportation. This costing should follow a "cradle to the grave philosophy" as indicated below:

(i) *Acquisition costs*
  (a) Research and development
  (b) System procurement
  (c) Non-recurring equipment

(ii) *Operating costs*

(iii) *Maintenance costs*

(a) Base
Materials - spares, aerospace ground equipment (AGE), etc.
People
Facilities
Training

(b) Depot
Materials - spares, AGE, AGE spares, etc.
People
Facilities
Training
Transportation.

The obvious question is where and how does the user secure these cost figures? The answer is that the cost-of-ownership rationale, if used by the user and supplier from program inception, will provide data and cost figures necessary for system selection. This can best be illustrated by considering the use of cost-of-ownership in inertial system design. We shall do this in part, by outlining our own history and experience and the cost-of-ownership techniques we have developed.

4. COST-OF-OWNERSHIP - PERFORMANCE IN DESIGN

In 1963 the Air Force Systems Command expressed deep concern over the rising costs and low reliability of Aircraft Inertial Navigation Systems. Those developing and delivering inertial navigation equipment have had to increase performance, due in part to higher performance aircraft and more severe operational requirements. The accomplishment of this, actually or theoretically, has required designs to be advanced to the very brink of the state of the art. In some areas, this increased risk and complexity has resulted in reduced reliability. The Air Force concern has had a multitude of results, one of which was the establishment of Detachment #1, Air Force Avionics Laboratory, at Holloman Air Force Base, New Mexico, in July 1963. This organization was established for two major reasons:

(i) To accomplish the in-house design, development, fabrication and test of a low cost inertial navigation system which utilizes the best components available without being restricted by the walls of proprietary bias.

(ii) To increase substantially Air Force in-house capability in the inertial navigation and guidance area[2].

There are two main ideas in (i). First, the equipments from different manufacturers are used and integrated into an operable system. Second, the approach is to low cost-of-ownership, not merely low cost. Neither of the foregoing has been accomplished previously for inertial navigation equipment used for military aircraft. In order to select the best components, the Detachment embarked on a subsystem basis as the only reasonable approach leading to a correct selection. Why subsystems? Consider the following:

Mr Magellan requires gyroscopes for his inertial navigation system. He listens to various manufacturers extol the advantages of their particular approach and design. As a result of discussions with many manufacturers, Mr Magellan selects three gyros whose characteristics are summarized in Table IV. He quickly selects Gyro B by reasoning that its cost per MTBF hour is $5.0/hr and the cost for two gyros is $8000, as opposed to $10,000 for Gyro A and $10,500 for three of Gyro C. Further investigation uncovers the need in Gyros A and B for a redundant axis capture loop which Mr Magellan determines will cost $700 with an MTBF of 6000 hours. Table V is a revised cost effective table showing more of a total approach.

TABLE IV

Gyro

	A	B	C
Type	Two-degree-of-freedom	Two-degree-of-freedom	Single-degree-of-freedom
Performance	0.01°/hr	0.01°/hr	0.01°/hr
Cost ($)	5000	4000	3500
MTBF (hours)	1000	800	700
Cost/MTBF	5.0	5.0	5.0

TABLE V

	MTBF (hours)		
	Gyro A	*Gyro B*	*Gyro C*
Gyro #1	858	705	700
Gyro #2	1000	800	700
Gyro #3	----	---	700
Gyro "subsystem"	460	375	254
	COST ($)		
	Gyro A	*Gyro B*	*Gyro C*
Gyro #1	5000	4000	3500
Gyro #2	5000	4000	3500
Gyro #3	----	----	3500
Redundant axis	700	700	----
Total	10,700	8,700	10,500
Cost/MTBF ($/hr)	23.1	23.1	45.0

Next he realizes that the output signal from Gyro B contains high and undesirable signals at particular frequencies, and that various filters must be added to his servo loops at an estimated cost of $90/servo loop for a total of $270, exclusive of assembly time. The basic servo loops for Gyros A and B are estimated to have an MTBF of 4500 hours which, with the addition of the notch filters, becomes 3200 hours. The estimated costs for the servo loops are $3500 without filters and $3770 with filters. Similar values for Gyro C requirements were estimated to be a cost of $2750 with an MTBF of 5000 hours. Another revision shows the values in Table VI.

TABLE VI

	Gyro A	*Gyro B*	*Gyro C*
Cost/MTBF	33.9	37.3	59.3

Additional thought on Mr Magellan's part coupled with subsequent tests by manufacturers A and B provides the data of Table VII.

TABLE VII

	Gyro A	*Gyro B*
Temperature sensitivity	$0.1^\circ/\mathrm{hr}^\circ\mathrm{F}$	$0.01^\circ/\mathrm{hr}^\circ\mathrm{F}$

In order to maintain adequate system performance using Gyro A, a temperature control of better than 0.05°F is required in order to reduce this error source to 50% of the specification value. This is translated into a cost differential of $1000 and a lower MTBF. The cost/MTBF between Gyro A and Gyro B is reversed, and Mr Magellan reflects on his original decision. This line of reasoning can be continued to the level of detail necessary to determine which of several possible choices should be made. Naturally, the selection can be only as good as the input data but, nonetheless, an analysis based on valid assumptions is significantly better than none at all. Furthermore, if users apply essentially the same reasoning to equipment selection, manufacturers will have no alternative but to determine and supply the information as needed. Admittedly, the foregoing example is a simplification of the real problem at hand. However, its main purpose is to provide the user with the desire for a solution. All too often drift data, such as gyro rates, are questioned without equally important emphasis being placed on equally important parameters. The application of a subsystem philosophy precludes the inadvertant omission of these important parameters. In addition, this philosophy provides a base line by which components of widely varying characteristics can be combined and optimally selected.

As a further example of the trade-offs to be made in subsystem design, consider Table VIII, which is a partial listing of Unipolar Mechanization Errors[2] at the end of four hours flight (Initial latitude 45°, Velocity 1000 ft/sec, Northeast two hours, Southeast two hours):

The example of Table VIII serves to illustrate that at the component level design trade-offs can be similar in manner to a linear programming transportation problem: maximize the profit (performance) and minimize the cost-of-ownership. By using the 0.05% scale factor, the gyro performance could be adjusted to provide comparable system performance (0.01 noise, 0.1% scale factor) which, in reality, probably is the more proper approach. Obviously, the next step is to ascertain the cost and reliability differential between using a 0.01°/hr gyro as compared to 0.015°/hr and a 0.1% scale factor as opposed to a 0.05% design. With the inherent stability of magnetic materials, the 0.05% design is considered to exhibit a relatively minor impact on design difficulty, cost and reliability, as compared to the 0.1% design. However, since gyro assemblies represent a complex assembly of interconnected and inter-related precision parts, the 0.015°/hr component probably will represent a component of lower cost and greater reliability. Obviously, this process can and should be analyzed more adequately for the entire ensemble of errors within each subsystem. All too often, equipment has been procured on the basis of what is available rather than what can be manufactured on the basis of sound technical and cost-of-ownership trade-offs.

TABLE VIII

Error Source		*Position Error*	
		R_x	R_y
X gyro noise (^{o}/hr)	0.01	2200	8700
(T_c = 1 hour)	0.015	3300	13050
Y gyro noise (^{o}/hr)	0.01	18000	4400
(T_c = 1 hour)	0.015	27000	6600
X gyro scale factor (%)	0.05	3250	14000
	0.1	6500	28000
Y gyro scale factor (%)	0.05	6500	6000
	0.1	13000	12000
Z gyro scale factor (%)	0.05	8000	500
	0.1	16000	1000

RSS error after four hours using components having 0.01^o/hr noise and 0.1% scale factor is 42,600 ft.
RSS error after four hours using components having 0.015^o/hr noise and 0.05% scale factor is 36,100 ft.

The preceding examples show how, by using a technical and cost-of-ownership trade-off rationale, the system designer can select and design the best components for each subsystem. Further, the system designer is accumulating cost-of-ownership data for each and every component and subsystem. In order to design a low cost-of-ownership system, quantitative data describing each component is necessary. Such detailed data, concerning cost, performance, reliability, maintenance, etc. can be gathered only if cost-of-ownership is considered in each step of the design process. The Detachment, in order to accomplish this, divided into four major subsystems as follows:

1. *Gyro Subsystem*
   (a) Gyros
   (b) Signal conditioning electronics
   (c) Stabilization and compensation electronics
   (d) Gyro torquing electronics
   (e) Redundant axis electronics

2. *Accelerometer Subsystem*
   (a) Accelerometers
   (b) Signal conditioning electronics
   (c) Rebalance loop electronics
   (d) A/D conversion electronics

3. *Gimbal Subsystem*
   (a) Gimbals
   (b) Platform torquers and resolvers
   (c) Resolver A/D conversion electronics
   (d) Wiring interconnections
   (e) Platform vibration isolators

4. *Computer Subsystems*

(a) Computer
(b) Display
(c) Control Panel
(d) Altimeter encoder

Each subsystem has three additional items which are common. These are

(i) Subsystem AGE
(ii) Subsystem non-recurring equipment
(iii) Ancillary equipment.

Once the subsystems listed previously are defined, the total cost-of-ownership of the system (CO_{sys}) is the summation of the cost-of-ownerships of each subsystems (CO_{-s}) plus miscellaneous costs associated with their integration. Therefore,

$$\text{Total } CO_{sys} = CO_{GS} + CO_{AS} + CO_{GIS} + CO_{CS} + C_M$$

where CO_{sys} is the cost-of-ownership of total system
CO_{GS} is cost-of-ownership of gyro subsystem
CO_{AS} is cost-of-ownership of accelerometer subsystem
CO_{GIS} is cost-of-ownership of gimbal subsystem
CO_{CS} is cost-of-ownership of computer subsystem
C_M miscellaneous costs required by integration.

Each of the foregoing subsystems has associated costs as shown on Figure 3. A detailed discussion of the subsystems developed by the Detachment is included in the Appendix.

5. COST-OF-OWNERSHIP - PERFORMANCE IN SYSTEM SELECTION

We now have indicated how a user can secure the basic selection data from suppliers by directing the use of performance and cost-of-ownership trade-offs during the design phase of the initial system. But it must be emphasized that this is only the basic data. The user must consider at least the following additional factors:

A. Use Rate, Pipeline Time and Repair Costs

Considering again our previous example of Systems A and B in Section 2, it is obvious that total costs are extremely sensitive to use rate and pipeline time. This sensitivity is so great that the user should concentrate his effort on increasing the effectiveness of the pipeline in addition to the hardware at this stage of the procurement cycle. This sensitivity and its importance is illustrated in Figures 4 and 5.

After reassessing the use rate, pipeline time and repair costs with the aid of appropriate Air Force repair facilities, the user can amplify or modify the contractor's cost-of-ownership data. Specifically, this will include the requirements for additional systems, and/or spare parts, which will have an end effect on cost-of-ownership.

B. Operating and Maintenance Procedures

Aviation Week states that "New data on avionics equipment failure rates indicate that the user and his maintenance procedures may have as great an effect on reliability as the equipment manufacturer".

Further, *Aviation Week* found that the same equipment which is subjected to different user operating and maintenance procedures can have variations in MTBF by a factor of as much as 10:1. It is imperative that acceptable procedures be defined at the outset (or at some milestone date) and attendant costs be assigned. With such information, the user will be able to assess his operating costs per hour and to identify those cost items in the contractor's proposal that require modification. In addition, some repair philosophy has to be established so that an economical break-even point is determined. This, in effect, stipulates an acceptable cost to be incurred in the repair of a piece of equipment. Newark Air Force Station has determined this to be 60% of the original system procurement cost for any one repair. Therefore, in the total cost history one must account for this funding requirement.

C. Inventory Control

The cost-of-ownership should include those items previously outlined in addition to inventory management as defined by the Logistics Management Institute[3]. LMI states that "estimates of the annual cost of holding an item in inventory range from 15 - 25% of the average inventory value". While this statement is significant in itself, one readily can see that now the basic maintenance concept must be defined with regard to a repair or throw-away concept.

We have already accumulated the data on pipeline time, use rate and repair costs. Using these, the user can generate more specific cost factors and draw a comparison between two basic approaches: repairable and throw-away.

Using the repairable approach the user must of course consider all costs: acquisition, operating and maintenance, both base and depot. Using the throw-away concept, the user considers all costs except depot maintenance but he must add further costs for spares.

The consideration of these two basic approaches is considered significant from two standpoints:

1. Will the throw-away concept have less effect on system MTBF as a function of the number of repair cycles?
2. Will the reduction in pipeline equipment requirements offset the additional base inventory required?

Now that the contractor's or suppliers' cost-of-ownership data has been amplified and modified, a comparative tool is necessary to complete the selection process. Since the parameters of greatest concern are cost-of-ownership and performance, these parameters can be considered to be factors of a product and are defined as follows:

$$\text{COPE} = \text{Cost-of-Ownership} \times \text{Performance (Ref.4)} \; .$$

CO now is defined as the total cost of the system in dollars per hour of operational life. System PE is defined as the accuracy of position indication in nautical miles per hour (3σ radial) for a selected flight path and duration. That system which exhibits the lowest COPE number for a given performance or for a given cost-of-ownership represents the best selection. Unless applied to a specific requirement, either performance or cost, the lowest overall COPE number would provide a system probably either too expensive to operate or performance much better than needed. Regardless of how the cost categories are grouped, they must include all cost items such that their total is the total cost-of-ownership of the system. The user could:

(i) For a given cost, select the system that maximizes performance.

(ii) For a given performance, select the system that minimizes the cost-of-ownership.

But why not select the development concept that will minimize the COPE product which is suitably weighted with cost-of-ownership and performance factors? These weighting factors may be considered, for example, as

$$\mathrm{COPE} = \mathrm{CO}_{TS}[W_1(P_{\mathrm{ratio}\ 1}) + W_2(P_{\mathrm{ratio}\ 2}) + \cdots + W_j(P_{\mathrm{ratio}\ j})] \ ,$$

where CO_{TS} is the cost-of-ownership of the total system P_{ratio} is a ratio of desired performance ($P_{j\ \mathrm{desired}}$) to contractor proposed performance (P_j) which is always equal to or greater than 1.

i.e. if desired performance is better than that proposed by contractor

$$P_{\mathrm{ratio}\ j} = P_j/P_{j\ \mathrm{desired}}$$

and if contractor performance is better than that desired then

$$P_{\mathrm{ratio}\ j} = P_{j\ \mathrm{desired}}/P_j$$

$j = 1$ to total number of performance factors considered, for example, 1 = position error, 2 = velocity error and 3 = azimuth error, etc.

w_j = weighting factors determined by mission requirements.

The cost-of-ownership performance product, as applied to an inertial navigation system, provides the rather unusual units of dollars mi/hr^2 or, for the weighted case (normalized), dollars per hour. The units of fiscal acceleration while unusual, should not present any insurmountable problem. These units would apply also to the various subsystems, since their particular errors would be related to equivalent mi/hr. We recognize that, in all probability, the information required to generate weighted COPE numbers is extremely limited. Perhaps it is not available. Nonetheless, this should not deter us from providing means for future data accumulation so that the most effective product selection can be made by the user.

6. CONCLUSIONS AND RECOMMENDATIONS

We have discussed the need for the application of a cost-of-ownership approach to inertial navigation system procurement and future development. A great deal of the information necessary to generate cost-of-ownership criteria is not available; however, the user, better than any other, can and should provide manufacturer motivation. It is not important that the approach discussed herein be adopted; however, some approach should be focused on navigation equipment from a total cost standpoint. This undoubtedly is necessary and action, perhaps through the medium of a government-industry committee or an organization such as AGARD, should be taken immediately to establish acceptable definitions, approaches and criteria.

REFERENCES

1. Weaver and Reichel — *The Air Force Design and Engineering Approach to Low-Cost Navigation Systems.* Navigation (Journal of the Institute of Navigation), Spring 1964.

2. *Unipolar Mechanization for Inertial Navigation.* Final Engineering Report. Autonetics, Division of North American Aviation. Contract AF 29(600)-4722.

3. *Life Cycle Costing in Equipment Procurement.* Logistics Management Institute, Washington, D.C. DDC AD 619-871, April 1965.

4. Laubendorfer, W.J. et al. *A LOw-Cost AircrafT Inertial NaviGation System - The Locating System*, AIAA/ION Guidance and Control Conference.

APPENDIX

The Locating System

The history of our Detachment and detailed descriptions of Low Cost Aircraft Inertial Navigation System (LOCATING) development have been the subject of previous papers[1,4]. Therefore this appendix is included for completeness only, in order to provide the reader with a brief description of our system development efforts.

As a result of various laboratory tests, industrial surveys, proposal evaluations, and in-house design efforts, the Detachment has completed the design definition of two inertial navigation systems. These LOCATING Systems, an acronym for Low Cost AircrafT Inertial NaviGation Systems, have been designated Gyroflex/Sperry (G/S) and Gyroflex/Kearfott (G/K). The G/S System is in the final stages of fabrication and test and will commence developmental flight test during late August 1967. The G/K System will enter this phase 8 - 10 months later.

The goals of each system are essentially the same, to minimize the cost-of-ownership for a given performance. The cost-of-ownership goal is $10/hour over the life of the equipment with a performance characteristic of 3 n mi/hr (3 σ radial).

The G/S System designation arises from the use of Gyroflex gyro with a Sperry gimbal set. A listing of the components and their suppliers included in each of the four major subsystems is as follows:

1. *Gyro Subsystem*

a. Gyro - Gyroflex
b. Signal conditioning electronics - General Precision, Detachment #1
c. Redundant axis electronics - General Precision
d. Gyro torquing electronics - Nortronics
e. Stabilization electronics - Detachment #1
f. Ancillary equipment (temp controller, power suppliers, etc.) Detachment #1

2. *Accelerometer Subsystem*

a. Accelerometers
(1) Horizontal - General Precision, Mod 2414
(2) Vertical - Autonetics, Mod A-40
b. Capture electronics and A/D
(1) Horizontal - Autonetics, Detachment #1
(2) Vertical - included in A-40 case
c. Ancillary equipment - Detachment #1

3. *Gimbal Subsystem*

a. Gimbals - Sperry, Detachment #1
b. Platform A/D - Ditran, Division of Clifton Precision Products Company, Detachment #1
c. Ancillary equipment - Detachment #1

4. *Computer Subsystem*

a. Computer - Nortronics
b. Control - Nortronics
c. Display - Nortronics

The interface engineering required to integrate the foregoing equipments has been completed in-house. In addition, other in-house accomplishments of significance are

(i) Design and manufacture of azimuth cluster.
(ii) Complete re-wiring of gimbal set to system requirements.
(iii) Programming of system mechanization equations.
(iv) Design and fabrication of all power supplies, various ancillary electronic equipment, power interrupt circuitry and system AGE.

The G/K System designation arises from the use of the Gyroflex gyro and the Kearfott ASN 58 (modified) gimbal set. In addition to the gimbal subsystem change, other major differences are

(i) All electronic boards have been redesigned to improve performance and reliability.
(ii) Vertical accelerometer - General Precision 2401.

This approach was taken since it appeared obvious that, with the Gyroflex and 2414, a new gimbal set would either have to be procured or designed in-house and manufactured on contract. Since a large portion of the developmental funding had already been applied by the government to the basic ASN 57/58 equipment, the decision was made to procure a modified version of the ASN 58 to Detachment #1 specifications. This approach provides a more effective utilization of in-house manpower and available funding.

The results of the laboratory and flight tests will be the subject of a technical paper when appropriate.

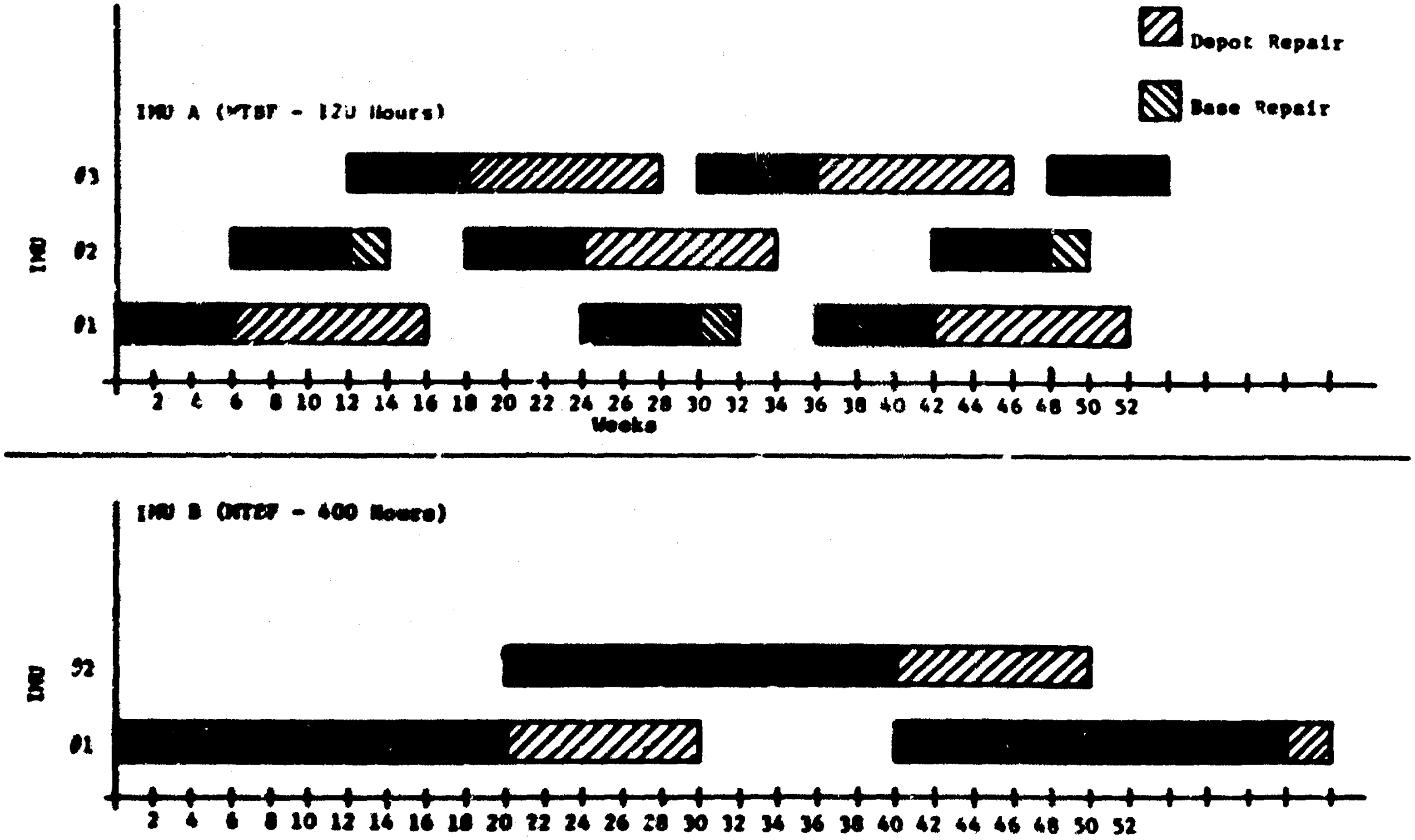


Fig.1 Operational and maintenance history of IMU A and IMU B. (Use rate - 20 hours/week, depot/base ratio - 2:1, base repair time - 2 weeks, depot repair cycle - 10 weeks)

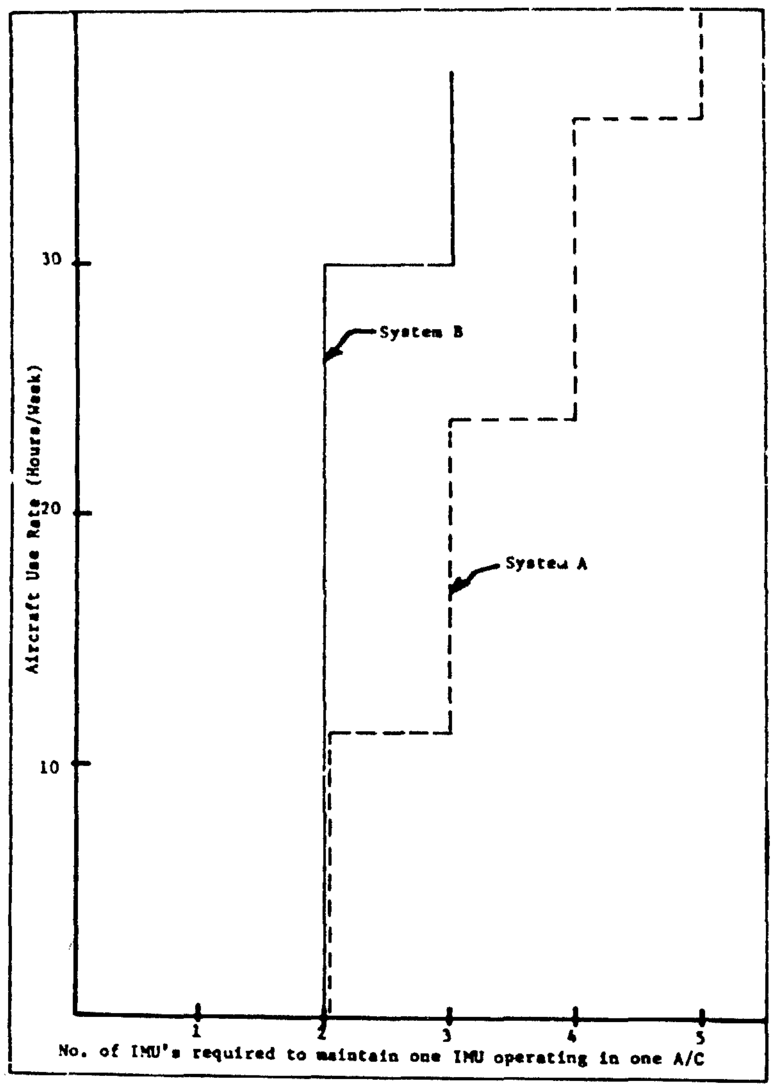


Fig.2 Number of systems required as a function of aircraft use rate

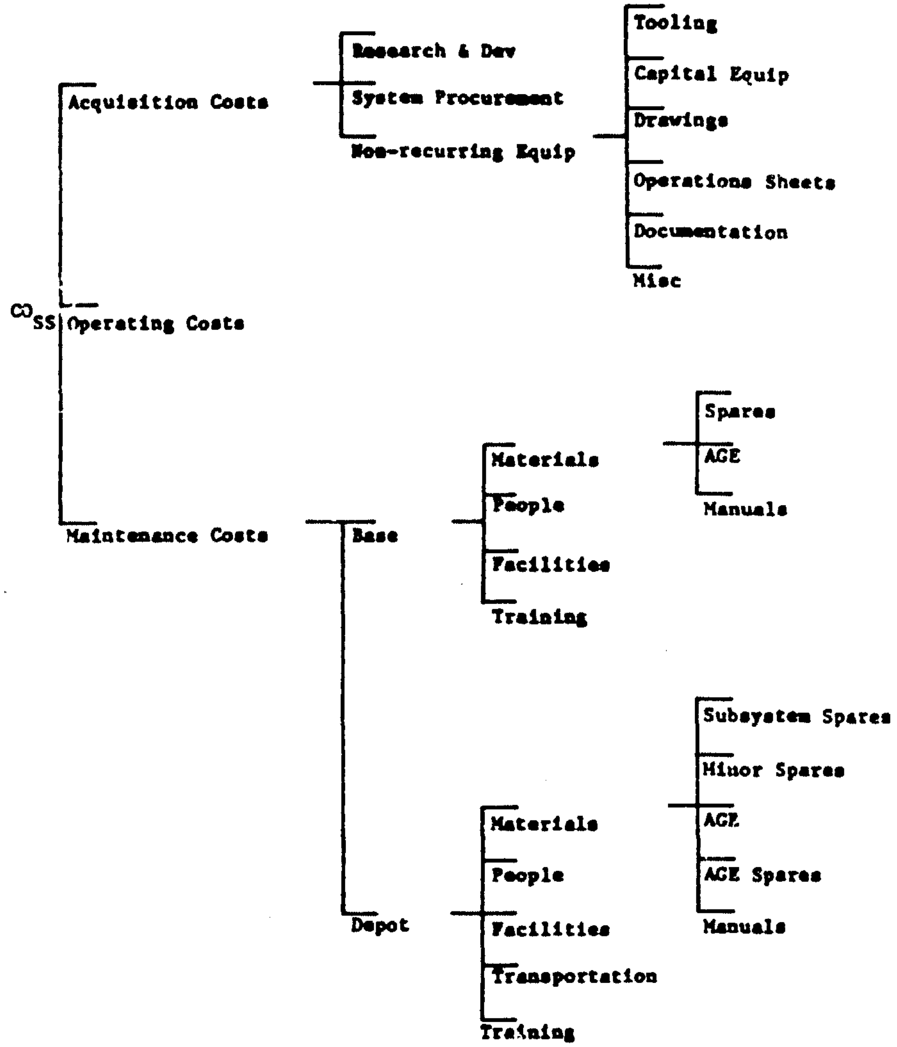
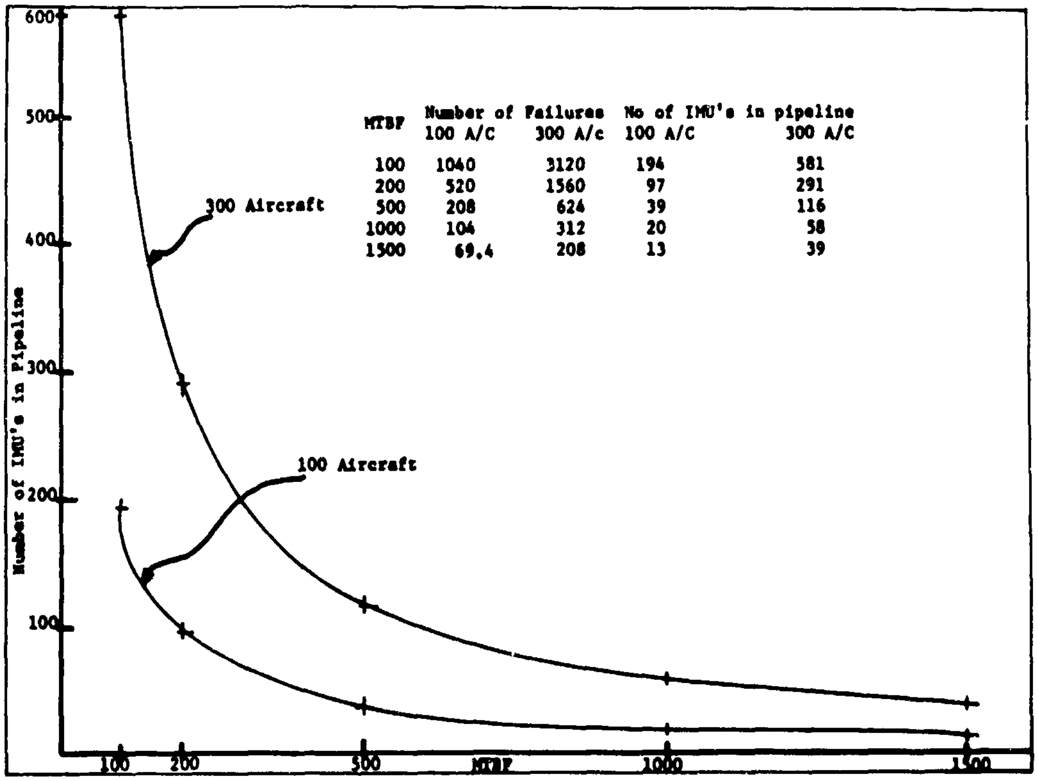


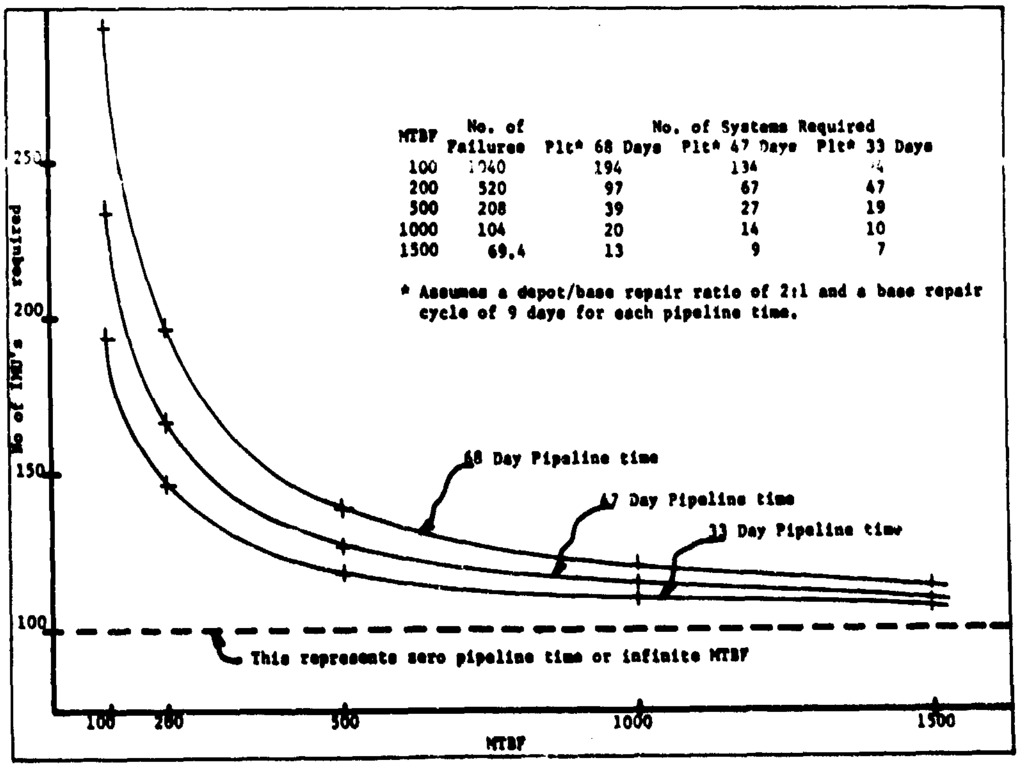
Fig.3 Subsystem costs

Best Available Copy



MTBF	Number of Failures 100 A/C	Number of Failures 300 A/c	No of IMU's in pipeline 100 A/C	No of IMU's in pipeline 300 A/C
100	1040	3120	194	581
200	520	1560	97	291
500	208	624	39	116
1000	104	312	20	58
1500	69.4	208	13	39

Fig.4 Number of IMU's in 68-day depot pipeline versus MTBF of IMU. Total IMU hours/year for 100 aircraft = 104,000, for 300 aircraft = 312,000. Use rate 20 hours/week per aircraft



MTBF	No. of Failures	No. of Systems Required Plt* 68 Days	Plt* 47 Days	Plt* 33 Days
100	1040	194	134	94
200	520	97	67	47
500	208	39	27	19
1000	104	20	14	10
1500	69.4	13	9	7

* Assumes a depot/base repair ratio of 2:1 and a base repair cycle of 9 days for each pipeline time.

Fig.5 Effect of pipeline time. Number of IMU's required for 100 aircraft versus MTBF of IMU. 100 aircraft, 20 hours/week (1040 hours/year) use rate. Total system hours/year for 100 aircraft = 104,000

A STUDY OF STRAPDOWN INERTIAL NAVIGATION AT THE MARSHALL SPACE FLIGHT CENTER

by

F.P.Daniel, G.B.Doane III, and R.R.Kissel

George C.Marshall Space Flight Center
National Aeronautics and Space Administration
Huntsville, Alabama, USA

SUMMARY

The history of inertial navigation contains a trend in which the hardware is becoming progressively more simple mechanically and at the same time more complex electrically. This trend is examined with particular emphasis on large rocket boosters. Attention is given to the generation of meaningful performance specifications. The system chosen at the Marshall Space Flight Center for detailed study and the reasons for its choice are discussed. Methods of statistical data reduction and some comparisons between testing components for a platform and for a strapdown system are presented.

NOTATION

a	rectilinear acceleration of the float along the i^{th} axis, referred to the float
D	drift (deg/hr)
H	angular momentum of wheel
IAF	input axis of float
I_{iiF}	moment of inertia of float about i^{th} float axis
I_{ijF}	product of inertia of float between i^{th} and j^{th} axes
I_{RA-RAW}	moment of inertia of wheel about its spin axis
$J_{\alpha f}$	polar moment of inertia of float about its input axis
$J_{\alpha g}$	polar moment of inertia of cradle about its input axis
K_a	anisoelastic drift coefficient
K_b	equivalent torsional spring rate of gas suspension bearing (cradle to float)
K_{bias}	g-independent drift rate coefficient
$K_{1,2,3}$	drift coefficient (deg/hr)/(deg/s)
$K_{4,5}$	drift coefficient $(deg/hr)/(deg/s)^2$
L_{IA}	projection of the distance from the float mass center to the flotation axis along the input axis of the float
L_{RA}	projection of the distance from the float mass center to the flotation axis along the spin axis of the float
m	mass of complete float assembly
M_c	constant drift term (deg/hr)
M_1	g-sensitive drift coefficient (deg/hr)/g
OAF	output axis of float
RAF	spin reference axis of float
T_α	net torque applied to the cradle input axis
T_β	torque applied to float about its output axis (usually around zero in servo stability analysis)
ω_I	inertial rate to gyro (table rate ± earth's rate (deg/s))
ω_N	natural frequencies of system
ω_S	angular velocity of the wheel with respect to the float

ω_{iF}	angular velocity of the float about its i^{th} axis, referred to the float
$\dot{\omega}$	time differential of ω
α	cradle input axis angle about the input axis
λ	added term in drift equation, second harmonic
β	float output axis angle (assumed signal generator input angle)
γ	added term in drift equation, second harmonic
θ	float input axis angle

A STUDY OF STRAPDOWN INERTIAL NAVIGATION AT THE MARSHALL SPACE FLIGHT CENTER

F.P.Daniel, G.B.Doane III, and R.R.Kissel

1. INTRODUCTION

1.1 General

As time has progressed, the application of gyroscopic navigation systems to vehicles has become more and more complex, both mechanically and electrically. Today, however, the new trend is that the mechanical features appear to be relatively simpler, even though systems are becoming more complex overall.

In this trend, which reduces the number of mechanical parts, it is helpful to examine the roles in which gyros have been applied. In doing this, one point becomes clear; i.e. gyroscopes have been applied in both a crude and a precision sense.

By way of illustration, one may cite many vehicles which used two-degree-of-freedom gyroscopes for basic attitude measurement and spring-restrained single-degree-of-freedom gyroscopes for attitude rate measurement. This application may be considered crude in the sense that the quality of the data derived from the gyroscopic instruments was not suitable for use in a self-contained inertial navigation system. In the past two or three decades, many systems have used gyroscopic attitude references to determine and control accelerometer orientation with respect to space on a stabilized, gimbaled platform. Typically, in this service, the attitude drift rate required for the precision navigation function is on the order of 0.5 deg/hr (2.4×10^{-3} mrad/s) down to such small claimed values as 0.001 deg/hr (4.8×10^{-6} mrad/s). By contrast, the rate gyros mentioned previously have accuracies on the order of 10 deg/hr (48×10^{-3} mrad/s).

The foregoing illustrates how equipment has become more sophisticated in its functions and thus complex in the overall sense. Most of the resulting systems use several gimbals carrying a stabilized inertial base on which, in turn, are mounted gyroscopes and accelerometers. Early systems had five gimbals; subsequent full inertial systems have used both three and four gimbal configurations. In general, the systems that do not have the extreme attitude requirements of aircraft have tended to use three gimbals. In the conventional platform inertial system, the stable element is either fixed in attitude with respect to inertial space or is slowly varying in attitude with respect to inertial space and fixed with respect to a moving navigational coordinate frame, such as an earth local vertical system. The gyroscopes used in this application are not required to sustain angular rates on the order of those experienced by the vehicle frame. Likewise, the accelerometers are not experiencing the severe angular motion environment of the vehicle frame. Indeed, there is still a rather large number of inertial system experts who believe that the gimbaled platform environment results in sufficiently improved sensor performance to justify the use of gimbals in any precision system.

Strapdown guidance is defined here as a system which consists of measuring instruments mounted to the vehicle frame; i.e. gyroscopes and accelerometers that provide information suitable for inertial navigation and the associated computer for interpreting this information[1]. In a strapdown system, the instruments are subjected to the total dynamic environment of the vehicle attitude. Additionally, the gyroscopes must provide information from which vehicle attitude will be determined. Because single-degree-of-freedom gyroscopes measure only angular rate, they must be followed by some form of integration procedure to determine attitude. The attitude data, when suitably combined with the accelerometer outputs, provide all the necessary navigation information. The mechanization concepts treated here are related to the inertial sensors and their operation. In particular for the attitude measurement, single-degree-of-freedom gyro sensors may be applied in two fundamentally different ways.

The first category is one in which a single-degree-of-freedom gyroscope is used to stabilize a "single axis platform" (SAP). A measurement of angular rate between the platform and the vehicle will provide the necessary measurement discussed previously. The second concept consists of single-degree-of-freedom gyroscopes, body-fixed to the vehicle. In this concept, a measurement of gyro output axis angle is used to hold the output axis angle at a very small value by applying a known torque to the output axis of the gyroscope. The value of torque is used as a measurement of vehicle angular rate.

1.2 Application

The application considered here is for the navigation of a large boost vehicle during mission phases including liftoff, ascent to earth orbit, earth orbit, and injection into a translunar or interplanetary trajectory. There are applications in which much time must be spent in earth orbit, perhaps as much as several weeks or even months. For this case, no known inertial systems are capable of achieving the low drift rates required. Therefore, realignment of all inertial systems is required after long duration in earth orbit or elsewhere. Thus, this discussion is restricted to ascent from the earth and injection into an earth orbit, coast in earth orbit for several orbits (e.g. four to ten), and subsequent injection into a transfer trajectory. For this case, it is feasible to use full inertial systems, unaided by realignment, for navigation to final injection. Such systems, augmented by a suitable realignment scheme, would be appropriate for the longer duration missions. The mission used to generate performance specifications is typical of many current and projected boost missions.

The figure of merit used to characterize the quality of an inertial system in the foregoing application is the accuracy with which a low circular orbit is achieved. The underlying reasons for choosing this criterion are as follows: From a propellant efficiency point of view, a nominally low altitude parking orbit is desirable; whereas from orbit lifetime considerations, a high orbit altitude is desirable. In general, therefore, we desire the lowest nominal altitude which has sufficient life for mission accomplishment when allowances are made for orbit uncertainties.

A simplifying assumption may be made to eliminate the absolute nominal altitude from further consideration. Briefly stated, the percentage loss in orbit lifetime caused by a perigee altitude error is proportional to this altitude error. For a typical configuration studied recently, this ratio is 10% loss in lifetime for 3500 meters altitude error at perigee. A safety factor of 1½ is used to account for additional uncertainties in orbit lifetime prediction and unaccountable items. The result is an allowable perigee altitude error of 5 km, which must be added to the ideal circular orbit altitude to obtain the nominal design altitude. When characterized *in toto* by just an angular drift rate, the required inertial system accuracy for a launch to a 185 km altitude parking orbit is 0.25 deg/hr (1.2×10^{-3} mrad/s). This value is a composite consisting of the equivalent average value of drift rate when all causes of error are considered. For a typical mix of error sources, this quality is adequate for final injection into the translunar or interplanetary orbit. Such accuracies are consistent with tracking accuracies and mid-course maneuver fuel requirements, i.e. the use of this criterion results in an inertial system that is adequate, but not overly so, for the entire mission. Table I gives the error sensitivities relating various inertial sensor errors to perigee altitude error. Figures 1 and 2 show perigee altitude error resulting from degradation of certain types of error sources.

In the strapdown application, it has been stated that one of the critical design parameters is angular rate of the vehicle body to which the inertial sensors are subjected. The total angular rate results from many causes; however, the following three categories provide a sufficient description: (i) The nominal slow, steady attitude rates commanded by the guidance system, (ii) rates caused by attitude maneuvers (resulting from changes in desired thrust direction, aerodynamically caused motions, bending, and sloshing), and (iii) random angular vibration. The characteristic values of these motions for a given system must be determined by analytic simulations and flight test measurements. Flight measurements are particularly important in this regard because none of the known vehicle simulations contain complete vehicle vibration models and little or no applicable, detailed flight data are available from the past. The most useful information currently available appears to be the telemetered signals from control rate gyros. These instruments seem most often to have a range of ±10 deg/s (0.17 rad/s) and respond to inputs of frequencies up to 30 or 40 Hz.

A frequency spectrum of some actual angular rates has been examined; typical results are shown in Figure 3. These analyses were run on worst case 10-second time slices. In general, they show a level of 0.0005 $(\mathrm{deg/s})^2/\mathrm{Hz}$ $[15.2 \times 10^{-8}\ (\mathrm{rad/s})^2/\mathrm{Hz}]$ with superimposed spikes between 1 and 6 Hz and between 35 and 40 Hz. The low frequency components correspond to the natural vibrational modes of the particular vehicle from which this information was obtained. It has been suggested that the high frequency components are caused by local mechanical conditions within the instrument itself, and that these could be brought down to 0.0005 $(\mathrm{deg/s})^2/\mathrm{Hz}$ $[15.2 \times 10^{-8}\ (\mathrm{rad/s})^2/\mathrm{Hz}]$ by suitable design of the control rate gyro mounting.

It is recognized that the upper limit of 40 Hz in the instrumentation may be restrictive, and instrumentation with wider bandwidth capabilities is being planned for some of the future flights. Other than to note that the strapdown instruments experience both translational and rotational vibrations, whereas a platform mounted component only experiences the translational, this nominal acceleration and velocity environment will not be discussed further because it is typical of large liquid-propellant rocket boosters.

Several flight tests have been made using single-degree-of-freedom gyroscopes in a strapdown configuration. Notably, the Ford Instrument Company's SAP and pendulous integrating gyro accelerometer (PIGA) concept was flight tested in an airplane in 1962.

Also, single-degree-of-freedom gyroscopes are being used in the precision rebalanced strapdown mode in the lunar module (LM) abort guidance system and the USAF PRIME/ASSET series of maneuverable reentry body tests. It is more reasonable to consider for the near future the use of single-degree-of-freedom gyroscopes as exemplified in these examples rather than other, more unusual types (such as laser gyros and magnetohydrodynamic gyros). Single-degree-of-freedom gyroscopes have been in use for a long time and much is known about them. Although this is a new application, many of the problems have been solved or are known to have solutions. Therefore, the mechanical and electrical problems associated with the application of single-degree-of-freedom gyroscopes are analyzed in some detail. A mathematical analysis of the mechanics and the servo of the single axis platform, as well as the mechanics of the precision rebalanced gyro, is included. Much of the analysis is the same for the two gyroscopes; however, the differences are indicated.

1.3 NASA/MSFC In-House Inertial Sensor Efforts

During the early in-house studies (about 1964), it was determined that MSFC could and should fill a gap which existed in the development of strapdown inertial hardware; the next step was to formulate an appropriate program. One of the ground rules decided upon was to make maximum use of applicable equipment stemming from current programs. Another ground rule required that the resulting system be competitive in performance with the best gimbaled platforms. Accordingly, the AB3 (3-cm wheel diameter) PIGA and the wheel and float of the AB5 gyro were postulated as components of an eventual inertial measuring unit (IMU). Thus, little change was necessary in a highly successful and available PIGA design; the only new design required was for the angular rate sensing device (SAP). To make a breadboard SAP available for laboratory testing in the shortest possible time, a cradle, with trunnion bearings, from a previous test program (Fig.4) was used. Its mounting surfaces were lapped to a tolerance of ±5 arc seconds.

The first system tests will only be of the IMU. The raw data from the sensors will be recovered for comparison with various known position time-line data. Later, however, full system tests will include a digital differential analyzer (DDA) type of computer which will operate on-line, processing each bit of data as it is generated by the SAP's or PIGA's and producing attitude and navigation (position and velocity) information. This computer is being designed and fabricated in-house at MSFC (as well as the IMU and the SAP's and PIGA's) and is reported upon in detail elsewhere[2].

2. ANALYTICAL CONSIDERATIONS

2.1 General

Although the gross theory of the single axis platform used for angular rate measurement is well known[3], the salient points will be reviewed. The desired information from the instrument is the rate at which the instrument base is rotating, with respect to inertial space, about one axis, which is usually designated the "instrument input axis". Rotations about the output and spin axes of the instrument should, ideally, not affect its indication. Considering an ideal instrument, it is readily seen that rates about the spin axis of the instrument will not be sensed because a gyroscope does not respond to rates about its spin axis. Rates about the instrument's output axis will, ideally, not be sensed either because, to keep the signal generator nulled, the input axis torquer will supply just enough torque to precess the gyro at the exact rate that the instrument is rotating about the float's output axis. The output from the instrument is obtained by measuring the angular rate between the instrument base and the gyro case. Unfortunately, errors can creep into any mechanization of this ideal operation.

2.2 Analog-to-Digital Conversion Considerations

The instrument's errors fall into several categories, one of which depends upon the choice of angular rate transducers. From a purely theoretical standpoint, an analog rate sensor (tachometer) would be ideal because the exact desired information would be continuously available. However, practical tachometers do not meet the accuracy requirements necessary for inertial navigation, and recourse must be made to measuring time and angular distance because these quantities may be most accurately measured. Since many transformation computer programs are available that can accept properly conditioned angular inputs, no fundamental limitation is imposed on the overall system if such a program is designed into the system from the outset. As a further practical matter, most modern guidance systems employ digital data processing, and one notes that the transformation computer will probably be amenable to operating on the discrete data resulting from the use of digital shaft angle encoders.

The remaining question then is what angular resolution is required. The answer depends on the accuracy sought from the overall system and hence is a figure distilled from many trade-off studies combined with knowledge of the technology of encoders, inertial sensors, and data processing equipment. (It would not be reasonable to install a 10 kg encoder on a 1 kg SAP, for example.) In this case, considering required system accuracy to be competitive (but not necessarily equal, in all respects) with a platform specifically designed for a high accuracy boost mission (roughly a 10-minute boost into a 185 km circular orbit), the required resolution, as determined by engineering simulations, is about 17 or 18 binary bits referred to 360 degrees (2π radians). Further refining will be done in future studies. Unfortunately, there currently appears to be a change in the class of shaft angle encoders available to achieve 18 bits (or more) as compared to 17 bits (or less); the dollar procurement cost is much greater for an 18-bit device. In addition, the requirements of the various trunnion bearings along the instrument's input axis become more stringent as more resolution is required. MSFC's first breadboard was built with a 16-bit resolution encoder, which happened to be available in-house from a previous encoder test program. Initial rate table testing soon showed the 16-bit resolution to be too low for determining the necessary performance data in a reasonable length of time (a day). Shortly thereafter a 19-bit encoder became available in-house for this program and was installed on the breadboard. The performance of this encoder was very satisfactory; the results, presented in Table II, were distilled from raw data obtained from the 19-bit encoder. A market survey of encoders suitable for a flight model of the SAP was conducted, and an 18-bit binary bit (resolution of 2π radians) incremental encoder is now being procured. These encoders must provide accurate information to calculate the rate data with tolerances on the order of the basic gyro drift rate, over a slew rate range of from 0 to ±1 rad/s. (Platform gimbal angle encoders only provide information for guidance command resolution.) The slew rate range, although somewhat arbitrary, since it must encompass rigid-body vehicle motions as well as local angular vibration rates, will include most practical large launch vehicle applications. Each of these two sources has been allotted approximately half of the total range. Other trade-offs are possible, depending principally upon the mission, vehicle, and vibration mount (if any).

2.3 Basic Dynamical Error Model

Basic instrument system errors may be treated in a fundamental, steady-state fashion by writing Euler's equations in which the angular rates are referenced to the gyro float. A rigid float in which an angular momentum vector is embedded provides a good model if none of the oscillatory modes associated with the contents of the float are excited. The IMU mounts must be designed to preclude this possibility, since large unpredictable drift rates would otherwise ensue. The torque summation equation about the float output axis is of particular interest here since it defines the gyro drift rate. This summation is as follows (see Notation):

Applied torque = inertial reaction torque

$$
\begin{aligned}
-mI_{RA}a_{IAF} + mL_{IA}a_{RAF} + K_a a_{IAF}a_{RAF} + K_{bias} = {} & \dot{\omega}_{OAF}I_{OA-OAF} + \\
& + \omega_{IAF}\omega_{RAF}\left(I_{IA-IAF} - I_{RA-RAF}\right) + I_{OA-RAF}\left(\omega_{IAF}\omega_{OAF} - \dot{\omega}_{RAF}\right) - \\
& - I_{IA-OAF}\left(\omega_{OAF}\omega_{RAF} + \dot{\omega}_{IAF}\right) + I_{IA-RAF}\left(\omega_{IAF}^2 - \omega_{RAF}^2\right) + \\
& + \dot{\omega}_{OAF}I_{OAOAW} - \omega_{IAF}I_{RA-RAW}\left(\omega_S + \omega_{RAF}\right) + \omega_{RAF}\omega_{IAF}I_{IAIAW} .
\end{aligned}
$$

Although well known, this equation still projects a rather formidable image. It may first be simplified by dropping all terms containing angular acceleration factors. One may also define, in the usual way, a quantity H which is called the wheel momentum (tacitly making the excellent assumption that $\omega_s >>> \omega_{RAF}$):

$$H = I_{RA-RAW}\,\omega_S .$$

The remaining inertial reaction terms are transferred to the left side of the equation. This procedure allows each "drift producing source" to be considered separately. Note that no viscous torque term is included in the equation; one might have been included for generality, but it is negligible for gas output axis suspension instruments such as the AB3 PIGA and the AB5 SAP.

From these analyses, one immediately notices that, if a gyro float were to be designed specifically for the strapdown application, the ideal moment of inertia tensor of the float would correspond to that of a sphere (all cross-products of inertia zero and all moments of inertia with respect to the OA-IA-RA triad equal). If this were accomplished, four terms would immediately drop out of the inertial reaction description. While some attention is given to "end-to-end" float balance in the course of designing gyros for use in three-axis stabilized platforms, there is no universally recognized requirement for dynamic (as opposed to the usual very careful static) balance. Some float designs have been proposed by others to achieve the spherical inertia tensor characteristic, but to our knowledge few have been built.

All the simplifications inherent in the output axis torqued case have been explored; however, in the SAP approach, all terms containing ω_{IA} automatically become negligible since the drift rate of the gyro about its IA will be very low (typically 0.5 deg/hr [0.25 × 10^{-3} mrad/s]). Thus, for the SAP, attention becomes focused on the terms $-I_{IAOAF}\omega_{OAF}$ and $-I_{IARAF}\omega^2_{RAF}$ as error sources in addition to those normally associated with inertial grade gyros for platform use. The only way to decrease these terms in a given environment is to reduce I_{IAOAF} and I_{IARAF} by redesign of the float or by sophisticated balance procedures, or both. Another phenomenon, which may or may not be immediately apparent from this equation, is the so-called output axis torque problem. It may be made visible by assuming that the servo, so far only tacitly a part of the instrument, has no long-term memory; i.e. the electronics transfer function is a constant at d.c. (the idea of using servos with memories in connection with gyros is an interesting separate research area which we have studied and designed into our latest SAP and PIGA servos). The terms of interest, which must balance, are

$$I_{OAOAF}\,\dot{\omega}_{OAF} = -\omega_{IAF}\,H\ .$$

If a constant angular acceleration about the output axis of the float occurred, the gyro would "drift" with respect to inertial space about its input axis at a rate given by

$$\omega_{IAF} = \frac{I_{OAOAF}}{H}\,\dot{\omega}_{OAF}\ .$$

This particular phenomenon has been reported[4].

2.4 Instrument Servo Design

Many servos for gas output axis suspension gyro instruments have been designed at MSFC; however, the problem was still of interest because, in designing such a large amount of angular momentum (approximately 2.5 × 10^6 g cm^2/s) into so small a package, the ratios of the various parameters are very different from those in any previous design. The analytical model actually used in-house for numerical design purposes is discussed here for the first time to the best of our knowledge; however, many more simplified versions are commonly known. Not nearly as detailed as the error model, the mathematical model used for servo and synthesis is developed from the following set of linearized (small angular motion) equations (see Notation):

(a) Torque equation assuming a torque applied to the cradle input axis:

$$T_\alpha = J_{\alpha g}\frac{d^2\alpha}{dt^2} + K_b(\alpha - \theta)\ .$$

(b) Torque equation about the float input axis:

$$K_b(\alpha - \theta) = J_{\alpha f}\frac{d^2\theta}{dt^2} + H\frac{d\beta}{dt}\ .$$

(c) Torque equation about the output axis of the float:

$$T_\beta = J_\beta\frac{d^2\beta}{dt^2} - H\frac{d\theta}{dt}\ .$$

The usual techniques of servo analysis yield the following transfer function for the mechanical portions of the loop.

$$\xrightarrow{T_\alpha} \boxed{\frac{1/H}{S\left(\frac{S^2}{W_{n_1}^{\,2}} + 1\right)\left(\frac{S^2}{W_{n_2}^{\,2}} + 1\right)}} \xrightarrow{\beta}$$

The output axis angle β is sensed by a signal generator, which produces an amplitude modulated voltage. A detector is used to convert the signal generator output to d.c. and to lower its impedance level. This signal, filtered to remove most of the ripple, is then used to drive a compensation network, which is mechanized by the use of operational amplifiers. The output from the compensation network is used as the input signal to a pulse-width power modulator which has low equivalent output impedance, reasonably linear input-output voltage characteristics, and small threshold values. A block diagram of this system is shown in Figure 5. A time constant is included in the description of the d.c. torquer because of the low output impedance of the pulse-width power modulator and the rather high natural frequencies of the SAP.

It is of interest to note that, if the torsional stiffness of the gas suspension bearing had been assumed infinite, as has been done in previously published accounts, only one natural frequency would appear in the SAP description; for this case, that frequency is on the order of 520 rad/s (83 Hz). This contrasts to the two natural frequencies in the description previously given (assuming a finite torsional stiffness), the lowest of which is 431 rad/s (68.5 Hz).

The compensation design must assure loop stability (all real parts of the roots of the characteristic polynomial corresponding to the system must have negative real parts) and achieve a rationally-based minimum loop stiffness (to assure physical maintenance of the instrument's float to case alignment). The actual form of the system's transient response is not too important as long as the settling time is short. A lower bound on the d.c. loop stiffness was established by assuming a cradle unbalance torque of 45 g cm in a 1g field and, hence, 450 g cm in a 10g field, which is the maximum design value of rectilinear acceleration used for Saturn. This requires a d.c. loop gain of 7×10^3 per second. Because the breadboard SAP had a much higher friction level than anticipated, both servo redesign and trunnion bearing preload adjustment were necessary. At this point, an active integrator was incorporated in the servo-loop electronics by using integrated operational amplifiers and a loop gain of $1.55 \times 10^4\ s^{-2}$ was achieved.

These overall requirements, combined with the necessity to avoid a conditionally stable loop in the linear mode, because conditional stability in the linear mode easily leads to limit cycle behavior during the initial nonlinear synchronizing mode, are sufficient to initiate a design cycle based on the classical root locus and Nyquist/Bode methods. Since these thoroughly practical methods require insight and intuition, tempered by as much experience as possible, we will merely exhibit the pole-zero plot (Fig.6) which resulted from the design iterations. This design results in a bandwidth around the loop of approximately 100 Hz. This bandwidth is required because of the high natural frequencies of the gyro. The electronic mechanization of the compensation used with the breadboard (which it is felt would probably be typical of flight systems) employs two commercially available operational amplifiers, although one of these is used to provide a telemetry pickoff point and therefore could, in principle, be eliminated. This design works well and yields a loop settling time of approximately 80 ms to a pulse disturbance (Fig.7). Further refining could always be done in designing future systems.

3. LABORATORY INVESTIGATIONS WITH BREADBOARD SAP

Upon completion of the breadboard, it was necessary to design an experiment to verify the SAP's inertial sensing behavior. The central piece of laboratory equipment used was a heavy duty, precision drive, rate table. The table is accurately oriented with respect to the earth's polar axis so that its precision rate driven axis is adjustable (by hand) in the north-vertical plane. As supplied by the manufacturer, this table was accurately instrumented only up to 200 earth's rate; however, instrumentation was added to make it accurate up to a table rate of 10 deg/s. This value, while still short of the design maximum of 1 rad/s. was at least only half an order of magnitude low. The next step in this type of testing is to assume some form of regression expression and then to determine if this expression adequately accounts for the gyro behavior. The expression chosen is (see Notation)

$$D = M_c + \mathrm{Mep}\ g_{OA} + \mathrm{Muby}\ g_{IA} + \mathrm{Mubz}\ g_{RA} + K_1\omega_{OA} + K_2\omega_{RA} + K_3\omega_I + K_4\omega_{OA}\omega_{RA} + K_5\omega_{RA}{}^2 .$$

The first four terms are those most often used in describing gyros intended for platform use (perhaps one should add the anisoelastic term, $E_0 g_{IA} g_{RA}$, as we will probably do in future SAP testing). The fifth, sixth, and seventh terms were included to account for and evaluate the overall alignment of the test set-up. The effects of these terms should not be charged against the intrinsic gyro or instrument accuracy because they may be made negligible by either mechanical or computational realignment, or both. The last two terms are those predicted for the SAP from the original look at Euler's equation applied to the gyro float.

Since MSFC's normal method of determining the M coefficients is to place the basic gyro in a planetary stand, isolated from earth's rate, no effort was made to design into the test set-up a really significant determination of these coefficients. Rather, attention was focused on determining the K coefficients, which evaluate phenomena unique to the SAP application. For these tests, the rate table rotation axis was permanently aligned parallel to the earth's polar axis. After some experimentation, 34 combinations of angular rate and gravity vectors, referred to the float, were chosen to yield the data for coefficient evaluation. These positions are tabulated in Table III; note that only in positions 33 and 34 does g_{IA} enter into the drift measurement (which is why these two positions appear). The other positions give combinations of values to the remaining terms. The high resolution (19-bit) encoder worked very well compared to a 16-bit encoder which was tried in the first test series. Test time in each position was only 7.2 minutes.

Typical results of the regression fits are tabulated in Table II. In general, the experimental accuracy improved with time. The most important error-causing factor was temperature. It was discovered that the first tests of this series had been started before the gimbal had temperature-stabilized in its thermal cover. When the cover was removed and the gimbal was exposed to room temperature only, the error was reduced by a factor of about four. Gas pressure was controlled to a nominal 103 N/m^2 (and readjusted each day on the same highly accurate gage to 103 N/m^2). Some g terms became significant and the misalignments were usually highly significant. The last two terms, however, were never significant at the 95% confidence level. The statistically significant values were usually very repeatable.

Tests were made in the given sequence most of the time; however, one set was run backwards and one was randomized, but the results were not much different from the other tests. The best tests have correlation coefficients of over 0.99 and standard errors of estimate less than 0.1 deg/hr. As is our standard practice, gyros were not recalibrated between runs nor was any output axis torquer used at any time during the tests.

These data represent very literally the first runs made on a new instrument being tested by new procedures for new phenomena. There has not yet been time for a long hard look at these data that are so necessary and productive in the long run. The program is progressing as expected and considerably more data will become available as additional SAP's finish the fabrication cycle and are tested in our inertial laboratory.

4. FUTURE PLANS

Future plans call for a two-phase experimental program backed by a continuing analytical effort. The first phase of the experimental program calls for the assembly and sled test of a complete IMU, without the transformation computer. Mock-ups of the IMU are shown in Figures 8 and 9. The outcome of this test series will provide both flight qualification testing of the IMU and raw instrument output data, which are desired for the computer development.

After completion of the sled program, it is planned that the IMU will be mated to its companion coordinate transformation computer and further rigorous testing will be pursued.

Some improvements in the instrument design details are anticipated but no major redesigns within the present experimental program are expected. Continuing development of engineering simulation, data recovery, and reduction software are programmed.

As an exercise in determining the growth potential of this concept with second generation instruments designed from the beginning for this application, a layout was made of an advanced IMU. The resulting mock-up is shown in Figure 10.

5. CONCLUSIONS

The work accomplished under MSFC's direction since 1964 has established that there are no analytical reasons barring booster navigation utilizing SAP's and PIGA's as strapdown inertial instruments. Simulations have demonstrated that, for injection into a 185 km circular orbit, accuracies competitive with platform derived navigation data are attainable.

It has also been realized that actual flight tests are needed to completely validate the assumptions used in the analysis. One of the most important areas in which it is necessary to gather flight data is that of the rotational vibration environment. These data are needed for further investigations into its effect on the instruments.

MSFC's choice of instruments has been shown to be a good choice for booster navigation. The results obtained from the MSFC developed methods of testing the SAP's and PIGA's, the accompanying error models, and the numerical results of the reduction do not, to date, exhibit any anomalies.

ACKNOWLEDGEMENTS

We wish to acknowledge the management support given the project by Dr W.Haeussermann, Director of the Astrionics Laboratory, and Mr C.H.Mandel, Chief of the Inertial Sensors and Stabilizers Division of the Astrionics Laboratory. We also wish to acknowledge the exhaustive and indispensable efforts of our co-workers. Although it is impossible to mention all the contributors, particular notice is due Messrs W.Panzer, I.T.Morgan, R.H. Tuggle, C.S.Jones, Jr., S.L.O'Hanian, C.J.Doran, E.Pikes, B.Walls, F.Nola and E.Ziesmer.

REFERENCES

1. Kennel, Hans F. — *Initial Alignment of a Strapdown Inertial Reference and Navigation System.* AIAA Preprint 67-556, August 1967.

2. White, J.B.; Acquith, R.C. — *A Coordinate Transformation Matrix Computer for Body-Fixed Reference Systems.* Marshall Space Flight Center, Huntsville, Ala. NASA TM X-53, 611, May 19, 1967.

3. Bessen, Alan — *Development, Design and Fabrication of a Pure Integral No-Gimbal System Breadboard Model.* Ford Instrument Co., Sept. 1961. (Unpublished)

4. Broxmeyer, Charles; Wishner, Howard — *Vehicle-Fixed Component Inertial Guidance System Study.* Final Report FR-4-359, Raytheon Co., July 14, 1964.

BIBLIOGRAPHY

1. Bowden, Roger — *Gimballess Attitude Reference System Study.* Convair, General Dynamics Corporation, December 31, 1958.

2. Wiener, Thomas Freud — *Theoretical Analysis of Gimballess Inertial Reference Equipment Using Delta-Modulated Instruments.* Instrumentation Laboratory, Massachusetts Institute of Technology, March 1962.

3. Bumstead, Robert M.; Vander Velde, Wallace E. — *Navigation and Guidance Systems Employing a Gimballess IMU. Guidance and Control - II.* Academic Press, August 12-14, 1965.

4. Powell, Jerry C. — *The Present and Future Roles of Strapped-Down Inertial Systems.* Transactions of the Eighth Symposium on Ballistic Missile and Space Technology, Minneapolis-Honeywell Regulator Company, October 16-18, 1963.

5. Gentry, C.D.; Johnson, F.V. — *A Bodybound Inertial Navigation System Using Electrostatically Supported Pendulous Gyro Accelerometers.* Transactions of the Eighth Symposium on Ballistic Missile and Space Technology, October 16-18, 1963, Wright-Patterson AFB, Ohio, and General Electric Company. (Unpublished).

6. Duncan, Robert Clifton — *1964 State of the Art, Navigation, Guidance, and Control.* NASA, Houston, Texas, AIAA Paper 64-500, June 29 - July 2, 1964.

7. Wing, W.G.; Macek, W.M. — *Current Status of the Ring Laser Gyro.* Sperry Rand Corporation, Unconventional Inertial Sensors Symposium, October 1964. (Unpublished).

8. Alongi, Robert E. — *Supplementary Analysis and Evaluation of Analytic Platform Systems - FY 64.* Report RG-TR-64-10, US Army Missile Command, July 10, 1964.

9. Langford, Robert C. — *Unconventional Inertial Sensors.* General Precision, Inc., AIAA Paper 65-401, July 26-29, 1965.

10. *No-Gimbal Feasibility Flight Test Program.* ASD-TDR-62-913, Navigation and Guidance Laboratory, Wright-Patterson AFB, Ohio, November 1962.

11. Haeussermann, Walter; Duncan, Robert Clifton. *Status of Guidance and Control Methods, Instrumentation, and Techniques as Applied in the Apollo Project.* AGARDograph 92, NASA, Marshall Space Flight Center, Huntsville, Alabama, and Manned Spacecraft Center, Houston, Texas. Presented at the Lecture Series on Orbit Optimization and Advanced Guidance Instrumentation, Düsseldorf, Germany, October 21-22, 1964.

12. Powell, Jerry et al. *The Gimballess Analytic Inertial Navigation System.* Minneapolis-Honeywell Regulator Company, May 31, 1962. (Unpublished).

13. *Gimballess Inertial Navigation Systems.* Moore School of Electrical Engineering; University of Pennsylvania, Semi-Annual Progress Report, November 1, 1964 - April 30, 1965.

14. Zagone, Peter V.; Hesse, Jr., Captain. *Inertial Attitude Sensors in a Strapped-Down System.* Holloman AFB, New Mexico, AIAA/ION Guidance and Control Conference, August 16-18, 1965.

15. Minor John W. *Low-Cost Strapdown Inertial Systems.* Honeywell, Inc., St. Petersburg, Florida, AIAA/ION Guidance and Control Conference, August 16-18, 1965.

16. Frasier, Cline W.; Gardiner, Robert A. *Lunar Excursion Module Mission and Related Guidance and Control Functions.* Manned Spacecraft Center, AIAA/ION Guidance and Control Conference, August 16-18, 1965.

17. Doennebrink, Fred; Russell, Jack. *LEM Stabilization and Control System.* Grumman Aircraft Engineering Corporation, AIAA/ION Guidance and Control Conference, August 16-18, 1965.

TABLE I

Error Sensitivities Relating Various Inertial Sensor Errors to Perigee Altitude Error. Mission Profile: Two Stage Ascent to Nominal Circular Orbit of 185 km Altitude

Error Source Name	*Error Source Unit Value*	*Perigee Altitude Error (meters)*
Roll axis accelerometer bias	10^{-5} g	180
Yaw axis accelerometer bias	10^{-5} g	96
Roll axis accelerometer scale factor	10^{-5} parts/part	295
Yaw axis accelerometer misalignment	1 arc sec	71
Initial system misalignment around pitch axis	1 arc sec	71
Constant drift rate around pitch axis	1 deg/hr	20,000
g-dependent drift of the pitch axis gyro	1 deg/hr g	31,000

TABLE III

SAP Test Input Conditions

(Test No. versus Acceleration and Angular Rate Input)

$D = M_c$	$+ \text{Mep } g_{OA}$	$+ \text{Muby } g_{IA}$	$+ \text{Mubz } g_{RA}$	$+ K_1\omega_{OA}$	$+ K_2\omega_{RA}$	$+ K_3\omega_I$	$+ K_4\omega_{OA}\omega_{RA}$	$+ K_5\omega_{RA}^2$
1	0	0	0.5686	0	-0.831	-0.831	0	0.69
2	0	0	0.5686	0	0.840	0.840	0	0.71
3	0	0	0.5686	0	-9.988	-9.988	0	99.76
4	0	0	0.5686	0	9.982	9.982	0	99.64
5	0.4021	0	0.4021	-0.588	-0.588	-0.831	0.35	0.35
6	0.4021	0	0.4021	0.594	0.594	0.840	0.35	0.35
7	0.4021	0	0.4021	-7.063	-7.063	-9.988	49.89	49.89
8	0.4021	0	0.4021	7.058	7.058	9.982	49.81	49.81
9	0.5686	0	0	-0.831	0	-0.831	0	0
10	0.5686	0	0	0.840	0	0.840	0	0
11	0.5686	0	0	-9.988	0	-9.988	0	0
12	0.5686	0	0	9.982	0	9.982	0	0
13	0.4021	0	-0.4021	-0.588	0.588	-0.831	-0.35	0.35
14	0.4021	0	-0.4021	0.594	-0.594	0.840	-0.35	0.35
15	0.4021	0	-0.4021	-7.063	7.063	-9.988	-49.89	49.89
16	0.4021	0	-0.4021	7.058	-7.058	9.982	-49.81	49.81
17	0	0	-0.5686	0	0.831	-0.831	0	0.69
18	0	0	-0.5686	0	-0.840	0.840	0	0.71
19	0	0	-0.5686	0	9.988	-9.988	0	99.76
20	0	0	-0.5686	0	-9.982	9.982	0	99.64
21	-0.4021	0	-0.4021	0.588	0.588	-0.831	0.35	0.35
22	-0.4021	0	-0.4021	-0.594	-0.594	0.840	0.35	0.35
23	-0.4021	0	-0.4021	7.063	7.063	-9.988	49.89	49.89
24	-0.4021	0	-0.4021	-7.058	-7.058	9.982	49.81	49.81
25	-0.5686	0	0	0.831	0	-0.831	0	0
26	-0.5686	0	0	-0.840	0	0.840	0	0
27	-0.5686	0	0	9.988	0	-9.988	0	0
28	-0.5686	0	0	-9.982	0	9.982	0	0
29	-0.4021	0	0.4021	0.588	-0.588	-0.831	-0.35	0.35
30	-0.4021	0	0.4021	-0.594	0.594	0.840	-0.35	0.35
31	-0.4021	0	0.4021	7.063	-7.063	-9.988	-49.89	49.89
32	-0.4021	0	0.4021	-7.058	7.058	9.982	-49.81	49.81
33	0	0.8227	0.5686	0	0	0	0	0
34	0	-0.8227	-0.5686	0	0	0	0	0

TABLE II

SAP Error Coefficients

TEST PARAMETER	SAP TEST DATA				PLANETARY DATA					
	JULY 29, 1966	AUG 1, 1966	AUG 1, 1966	AUG 2, 1966	FEB 28, 1966	MAR 15, 1966	MAR 16, 1966	MAR 17, 1966	JULY 26, 1966	AUG. 4, 1966
SEE	0.2042	0.4423	0.4517	0.3961	0.2179	0.1378	0.1587	0.1494	0.1926	0.0924
CORRELATION	0.9983	0.9966	0.9964	0.9980	0.9631	0.9589	0.9632	0.9471	0.9392	0.9946
M_c	-0.126 ± 0.073	-0.079 ± 0.157 †	-0.082 ± 0.159 †	-0.106 ± 0.141 †	-0.038 ± 0.382 †	0.146 ± 0.242 †	0.089 ± 0.279 †	0.074 ± 0.260 †	-0.059 ± 0.338 †	0.135 ± 0.162 †
Mop	0.016 ± 0.106 †	-0.083 ± 0.402 †	-0.084 ± 0.083 †	0.026 ± 0.360 †	-0.022 ± 0.664 †	0.018 ± 0.418 †	-0.012 ± 0.483 †	-0.032 ± 0.454 †	-0.007 ± 0.587 †	-0.094 ± 0.280 †
Muby	1.448 ± 0.309	1.324 ± 0.834	1.311 ± 0.851	1.527 ± 0.747	-1.768 ± 0.664	-1.046 ± 0.418	-1.299 ± 0.483	-1.014 ± 0.454	-1.194 ± 0.587	-1.825 ± 0.280
Mubz	0.856 ± 0.106	1.051 ± 0.401	1.069 ± 0.411	0.839 ± 0.360	-0.238 ± 0.664 †	0.188 ± 0.418 †	0.083 ± 0.483 †	0.051 ± 0.454 †	-0.260 ± 0.587 †	-0.750 ± 0.280
K_1	5.991 ± 0.149	6.245 ± 0.323	6.245 ± 0.323	6.204 ± 0.289	NOT APPLICABLE					
K_2	-0.722 ± 0.149	11.61 ± 0.323	11.44 ± 0.323	15.04 ± 0.289						
K_3	5.323 ± 0.105	5.560 ± 0.220	5.475 ± 0.233	5.350 ± 0.146						
K_4		FOUND INSIGNIFICANT EARLIER AND NOT COMPUTED HERE								
K_5										
γ	0.090 ± 0.149 †	-0.265 ± 0.323 †	-0.434 ± 0.330	-0.051 ± 0.289 †						
λ	0.105 ± 0.149	-0.662 ± 0.323	-0.662 ± 0.330	-0.532 ± 0.289						

NOTES:
1. THE ± FIGURE INDICATES THE PARAMETERS 95% CONFIDENCE LIMITS.
2. † DENOTES PARAMETER NOT STATISTICALLY SIGNIFICANT IN THIS TEST.

$$D = M_c + M_{op}\, g_{oa} + M_{uby}\, g_{ia} + M_{ubz}\, g_{sa} + K_1 \omega_{ia} + K_2 \omega_{ia} + K_3 \omega_I + K_4 \omega_{oa}\omega_{ia} + K_5 \omega_{oa}^2 + \gamma(\omega_{oa}^2 - \omega_{ia}^2)/\omega_I + \lambda \omega_{ia}\omega_{oa}/\omega_I$$

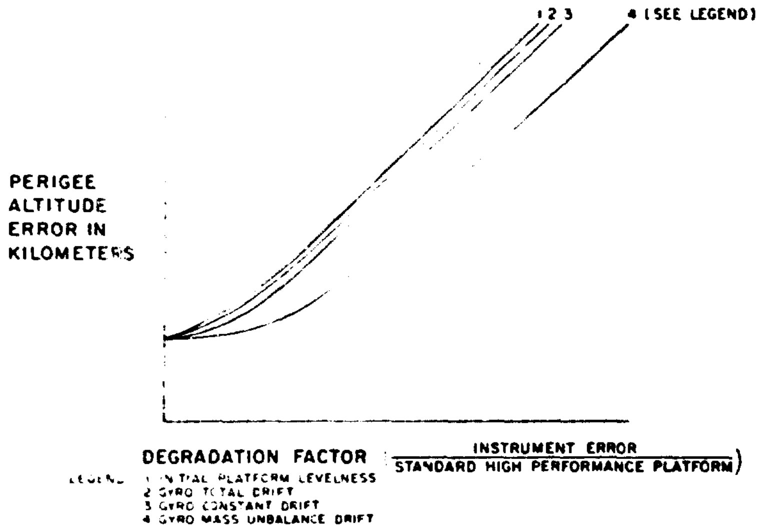


Fig.1 Perigee altitude error resulting from degradation of gyroscope and initial alignment accuracies. .(Based on typical state-of-the-art high performance systems)

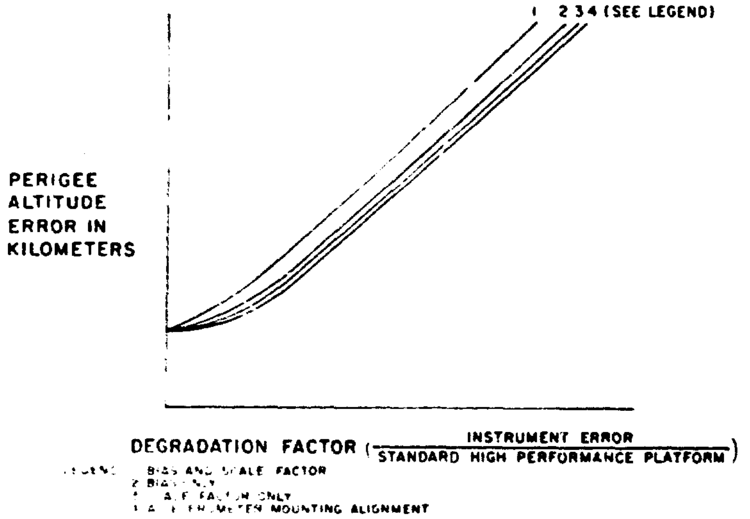


Fig.2 Perigee altitude error resulting from degradation of accelerometer accuracies. (Based on typical state-of-the-art high performance systems)

Best Available Copy

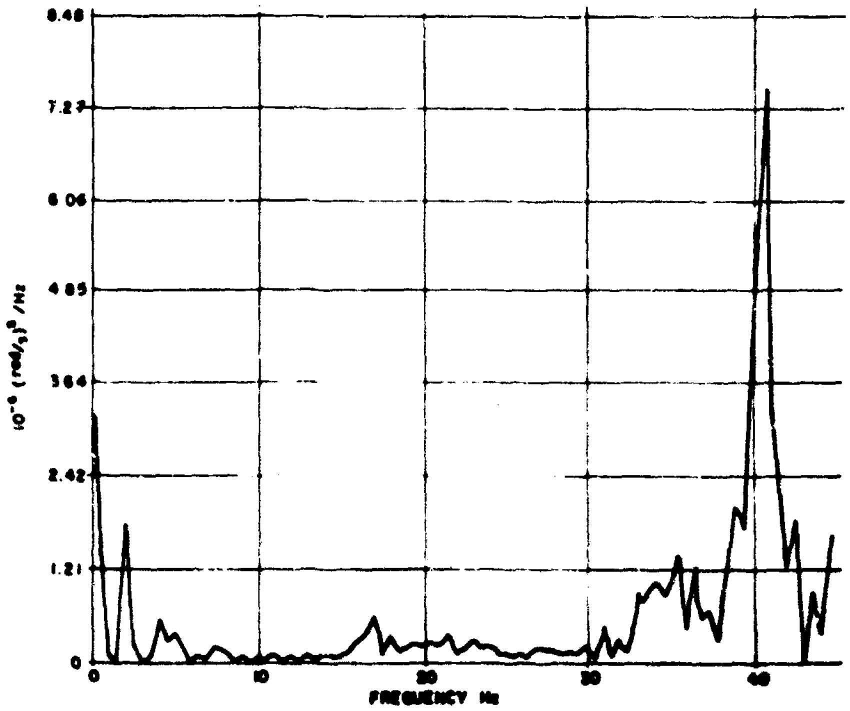


Fig.3 Representative spectrum of angular motion in Saturn instrument unit during flight

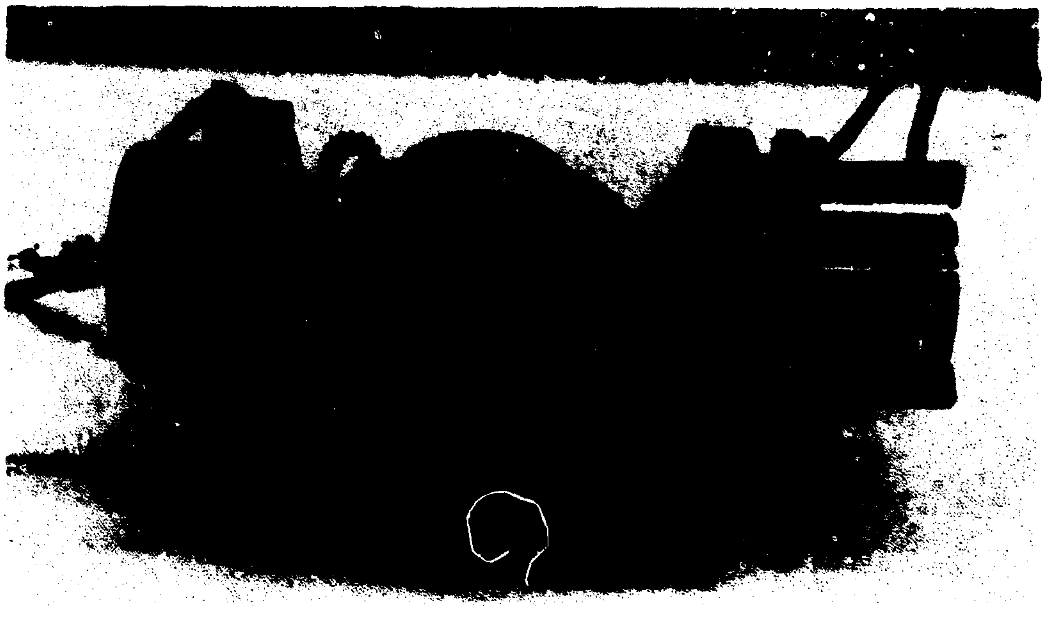


Fig.4 Breadboard SAP

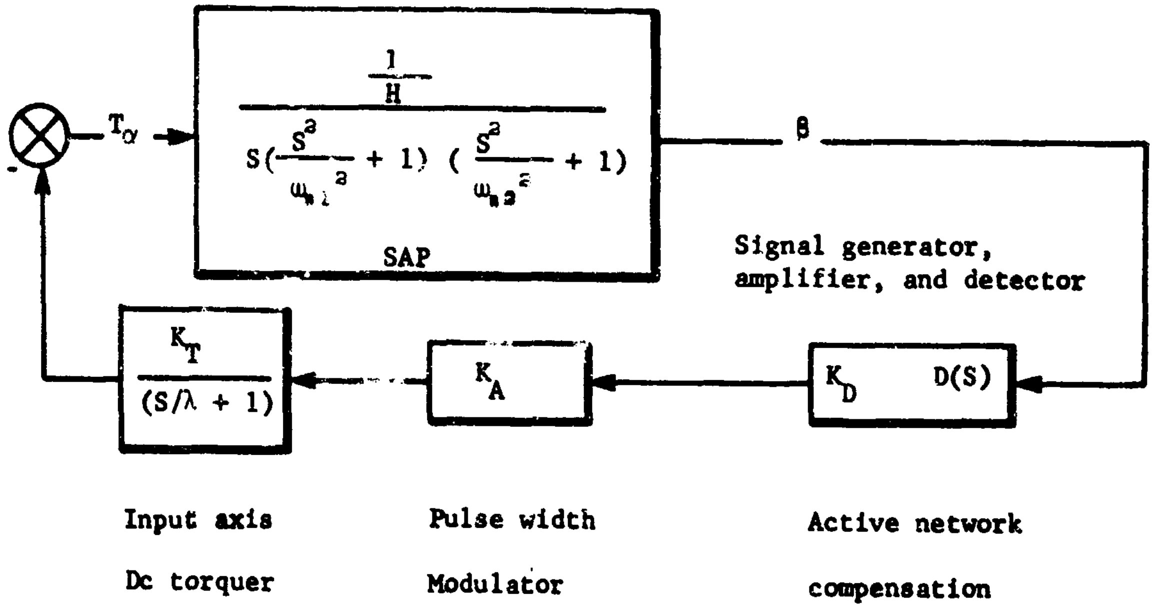

T_α
-
$\frac{1/H}{S\left(\frac{S^2}{\omega_{n1}^2}+1\right)\left(\frac{S^2}{\omega_{n2}^2}+1\right)}$
SAP
β
Signal generator, amplifier, and detector
K_D D(S)
K_A
$\frac{K_T}{(S/\lambda + 1)}$
Input axis
Dc torquer
Pulse width
Modulator
Active network
compensation

Fig. 5 Strapdown instrument servo

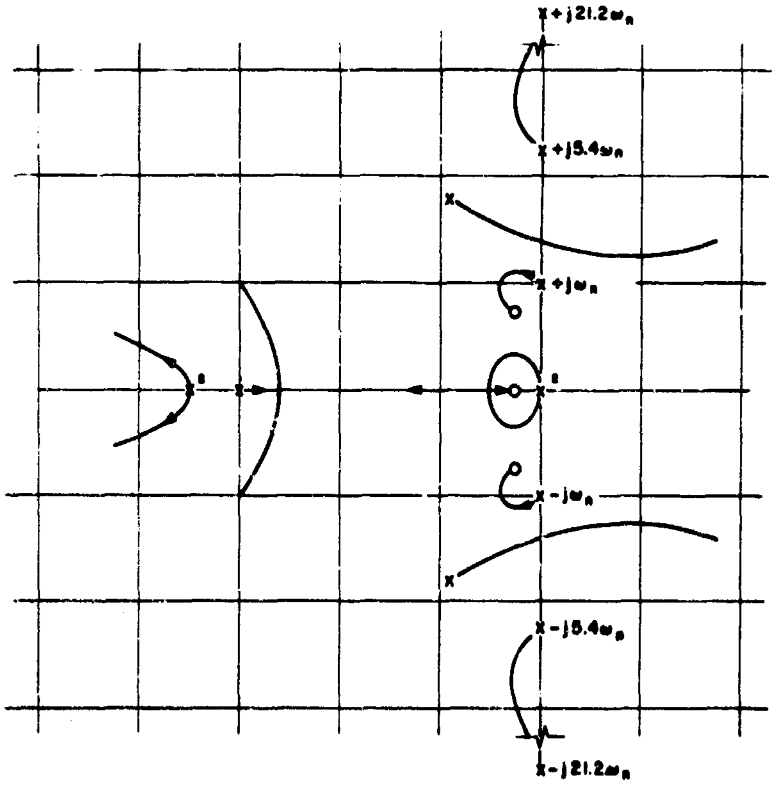


Fig. 6 Instrument servo root loci

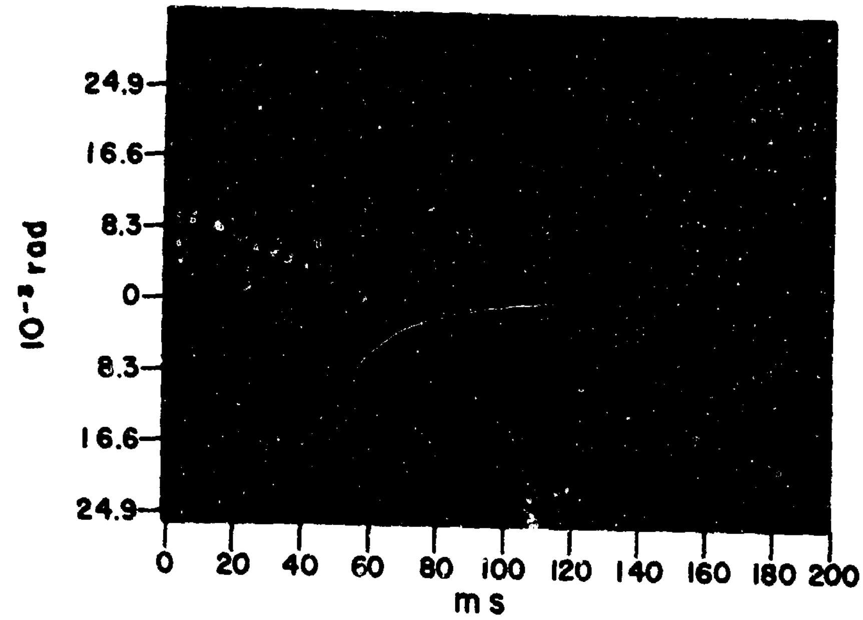


Fig.7 SAP servo transient response

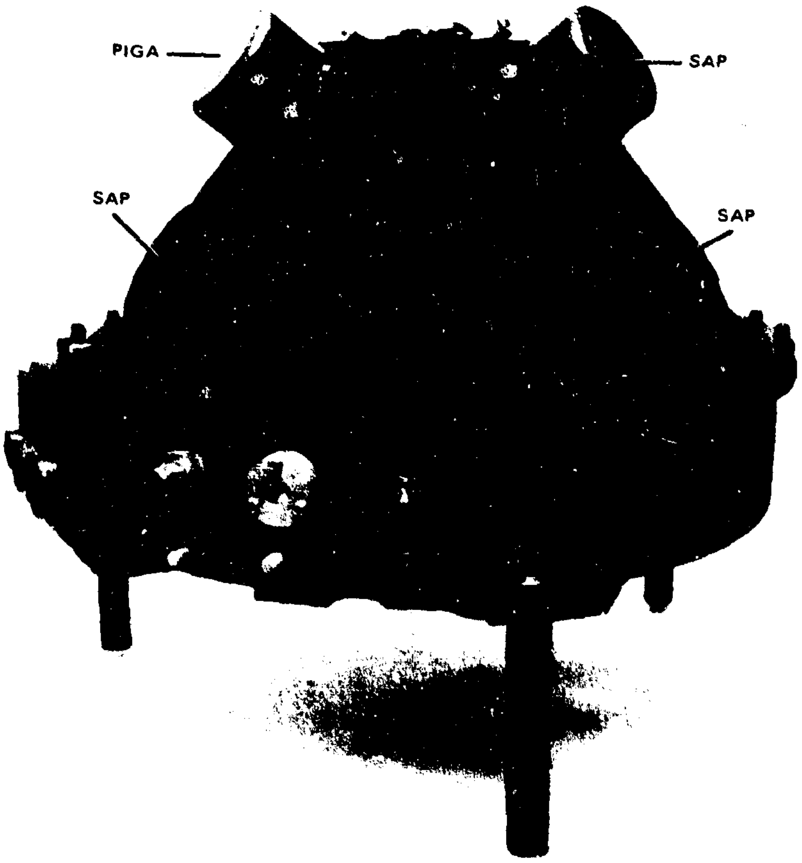


Fig.8 Strapdown platform, side view

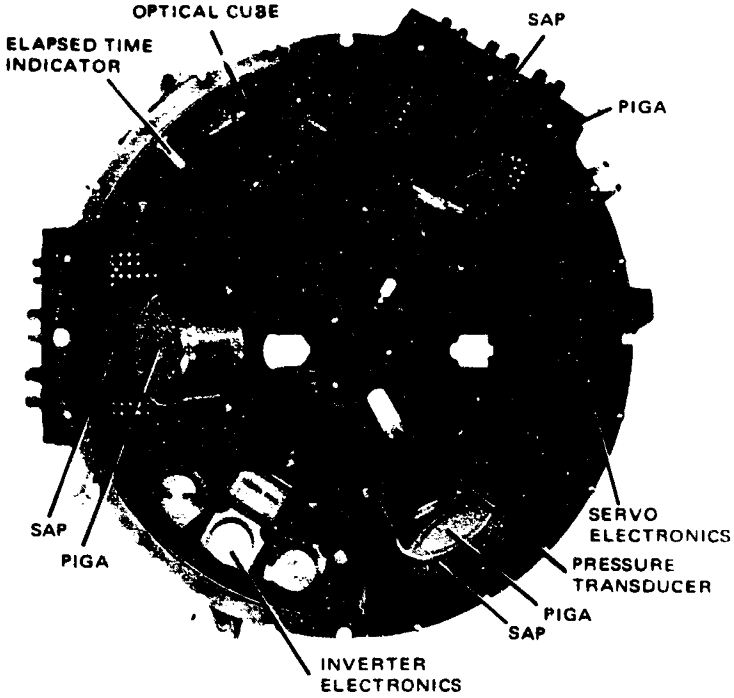


Fig.9 Strapdown platform, top view

Fig.10 Conceptual strapdown system

DESCRIPTION OF A STRAPDOWN INERTIAL MEASUREMENT UNIT

by

Joseph Yamron

United Aircraft Corporate Systems Center,
Farmington, Connecticut, USA

SUMMARY

A strapdown inertial reference package has been developed for space vehicles based on single-degree-of-freedom instrumentation. A number of these packages have been built for the Lunar Module's Abort Guidance System. Production-acceptance test data, which include both periodic calibration sequences and environmental tests and data acquired in connection with sled testing at Holloman Air Force Base, are reviewed. Error sources are examined for boost operations. Development problems encountered to date are highlighted and an assessment is made of strapdown applications.

DESCRIPTION OF A STRAPDOWN INERTIAL MEASUREMENT UNIT

Joseph Yamron

1. INTRODUCTION

For many years, gimballed platforms have been synonymous with inertial navigation. Although strapdown components have been used from the earliest days of aviation in control systems, the problem of continuously measuring high rates and electronically performing base motion isolation has been a formidable one indeed. The development of small, high speed digital computers reopened the possibility of producing a strapdown system. In 1959, United Aircraft Corporation undertook to examine the feasibility of building a high performance inertial guidance system using strapdown single-degree-of-freedom instruments. In 1963, under Air Force sponsorship, a system aimed at boost applications was successfully demonstrated in the laboratory and, in 1964, the development of the Abort Sensor Assembly for NASA's Lunar Module was begun by United Aircraft Corporation.

During the course of these efforts, a new test methodology and test equipment had to be developed. Analysis and simulation techniques were given extensive early attention, and these test and analytical tools were then brought to bear on the design of instruments and electronics to evolve the current system. Much of this work has been reported in the literature. Four complete generations of hardware have been produced and enough data accumulated so that this art can now be assessed. The long-held view that strapdown systems were impractical because of hidden dynamic effects on the one hand and impossible computation problems on the other was dispelled. The effects of temperature were found to be predictably critical, and long-term stability had to be traded off against the problems of calibration and alignment. The computer interfaces, in a vehicle system where the strapdown computation was only one of many tasks, had to be evaluated as part of this development program. This paper is concerned primarily with recent performance data on strapdown inertial packages which have resulted from this effort.

2. DISCUSSION OF THE PACKAGE DESIGN

2.1 Package Description

The strapdown inertial measurement unit (SDIMU) is a self-contained assembly designed to sense incremental displacements about the vehicle axes and velocity increments along the vehicle axes, and transmit these signals in the form of discrete pulses to a digital flight computer in order to perform the functions of:

(i) Calibration and/or alignment.

(ii) Attitude determination (base motion isolation transformation matrix).

(iii) Navigation.

(iv) Vehicle control.

(v) Guidance.

The SDIMU must accurately measure angular and linear motion while subjected to powered limit cycle and free fall environments during the course of its mission. For the Lunar Module Abort Sensor Assembly (LM/ASA), it has the following characteristics:

Size: 623 cubic inches (with mounting feet).

Weight: 20.7 pounds.

Steady-State Power (vacuum environment): 74 watts.

Operating Life: 5000 hours with maintenance
1000 hours without maintenance.

Reliability Goal: 0.9991.

Dynamic Range: 28 degrees/second angular rate
100 feet/second^2 linear acceleration.

Resolution: 2^{-16} radians/pulse ($\simeq$ 3 arc seconds) angular displacement
$\simeq$ 0.003 feet/second velocity increment per pulse.

The assembly as shown in Figure 1 consists of three UACSC RI-1139 floated, rate integrating, single-degree-of-freedom gyros; three Bell VIIB pendulous accelerometers and their associated pulse torquing electronics; a frequency count-down subassembly; a temperature controller unit; computer interface electronics; and an integral power supply.

Both the gyro and the accelerometer selected for the SDIMU operate on a torque balance principle. In the gyros, electromagnetic torquers are used to balance gyroscopic input torque. The average current in the torquers is equivalent to angular rate, and pulses (or time-modulated increments) of current are equivalent to, and used to indicate, angular increments. Similarly in the accelerometer, current is used to generate a torque that opposes pendulous torquers caused by acceleration (or gravitation). The average current level is proportional to acceleration, and pulses of current are counted to measure increments of velocity. All of the sensors contain torquers with permanent magnet stators for use with the pulse torque servo amplifiers (PTSA's).

The frequency countdown subassembly provides precise timing and reference signals for both the PTSA's and the gyro wheels. The temperature controller provides warmup and fine control; no instrument heaters are employed. The fine temperature controller utilizes two temperature sensors mechanically located in optimum sensing locations, electrically in diagonal legs of a bridge, to detect temperature variations within the ASA housing, and stabilizes its temperature by applying power to two heaters on the housing. The housing thermal conduction and gradient patterns are analyzed and synthesized in the mechanical design process to create for the sensors an optimum thermal environment with minimum complexity.

All electronics are microminiaturized, utilizing fully qualified integrated circuitry for minimum weight and maximum reliability. The packaging density that has been achieved is quite high for this type of equipment. The pulse torquing amplifiers provide sensor readout with an accuracy that approaches the sensor itself. These amplifiers furnish constant power to the sensor torquer to give the best possible sensor stability. The net current delivered to the sensor is a function of the input, while the power is independent of the input; equal negative and positive currents are delivered to the torquer when the inertial input is zero. The rebalance electronics include an output switching circuit that approaches the capability of mechanical switches (e.g. the ratio of open-circuit resistance to closed-circuit resistance is in the order of 10^7). The error caused by the circuit is related to the full scale torquing rate by this ratio. A system with a 28-degree/second (100,000 degree/hour) maximum input rate thus can expect to have an electronic circuit equivalent gyro drift rate of 0.01 degree/hour.

A feature of this design is that although the switching rate is held at 1 kc, an information rate of 64 kc is achieved. In the case of the accelerometer, increasing the switching rate will minimize vibropendulous errors. A switching rate of 1 kc was selected for both instruments, however, for simplicity.

The sensor loop implemented in the strapdown package is shown in Figure 2. It utilizes a form of time modulation to provide the rebalancing current for the sensor.

2.2 Key Development Problems

About 30,000 hours of testing were accomplished on several programs prior to the release of the production design for the LM/ASA. Since then an additional 70,000 hours have been accumulated on the production hardware. Early in 1967 a modified version of the ASA was tested on the rocket sled at Holloman Air Force Base.

The development of the gyro and its rebalance electronics had first priority. This mechanization has been treated by Scoville and Yamron[1] and further by Scoville[2]. The gyro design was intended for pulse torqued, high rate operation from its inception, and the principal requirement for this design was to provide excellent torquer stability in conjunction with high sensitivity.

The torquer stability was obtained by minimization of the sensitivities of torquer scale factor to axial, radial, and rotational motions of the float. Figure 3 depicts scale factor stability obtained on a group of 20 production gyros. A 1σ value of 2 ppm is evident. Figure 4 plots gyro/PTSA scale factor stability versus time for the three gyros in the slide-tested unit. The average 3σ value for a 5½ month period is 72.8 ppm.

One interesting problem was a nonlinearity caused by the magnetic time constant of the torquer, that is, a lag between the rise of current in the torquer and the rise of flux in the air gap retarded by eddy currents in the torquer frame. In the first model of the gyro the value of this time constant was over 250 milliseconds, which seriously compromised the linearity of the scale factor at rates over 0.33 radians/second. This was minimized by designing a slotted low resistance torquer frame which afforded a factor-of-three reduction. Figure 5 illustrates the excellent linearity achieved on an early gyro/PTSA channel.

A strapdown gyro must have a high damping coefficient in order to restrict float motion. Initially, this proved quite troublesome in the LM/ASA Program. A fluid with two components of different density had been chosen to meet certain low temperature aspects of the specification. To achieve high viscosity, a fluorocarbon was blended with a viscosity improver. Thermogravitational diffusion separation of the two components into discrete density layers resulted in mass unbalance torques. The drift rate associated with this effect is inversely proportional to viscosity. The major effect, however, takes place once separation has occurred and a given orientation of the instrument is established. A drift will be experienced whenever the gyro's orientation with respect to gravity is changed. The time required for this drift to settle out will be a function of viscosity and the difference in density of the fluid components. It may vary from several hours to a few weeks. When calibration is not practical for days or weeks before flight, however, the problem can become severe. By storing the gyros at high temperature (to accelerate the diffusion) prior to mass unbalance trim, and by using a preferred package orientation in the storage container, this problem is minimized. Incorporation of a narrow cut, single component fluid eliminates the difficulty altogether. To complete the argument, an experiment was conducted with a gyro using a viscosity improver of the same density as the fluorocarbon, and no such effect was observed.

A vexing problem peculiar to strapdown operation was uncovered during the early development testing of a complete system on a centrifuge. An apparent drift rate in the base motion matrix calculations during the dynamic rate environment occurred due to dissimilar bandwidth characteristics of the gyro channel loops. This effect was analyzed as a pseudo-coning drift rate in one channel as a result of phase mismatch in the other two channels when the latter experienced sinusoidal inputs. Figure 6 illustrates the relationship of the pseudo-coning drift as a function of channel phase mismatch and sinusoidal rate amplitudes. Because of this error source, all gyro channels are trimmed during in-line testing to be within a narrow band at low frequencies.

The RI-1139 gyro was optimized to control input axis stability by the use of 25-microinch clearance, pivot-jewel output axis bearings. Also, the equalization of the inertia of the gyro float for axes in the plane normal to the output axis was incorporated.

The accelerometer utilized in the SDIMU is the Bell Aerospace Corporation Model VIIB single-axis, pendulous, permanent-magnet, pulse-rebalanced unit. In pulse torquing operation, the torquer is operated at a voltage above ground which creates a difference of potential between the pendulum and the sensing capacitor pick-off plates. The torquer coil is wound on the pendulum. This results in a force of attraction which disturbs the initial trimmed "electrostatic null position" away from the mechanical null. The bias is perturbed until the charge is bled off each time the unit is energized. This problem was initially corrected through the use of trim capacitors. A superior method uses bleed resistors placed in the capacitance bridge circuit. These resistors decrease the bias sensitivity to a negligible value and improve turn-on time to reach steady-state outputs.

2.3 Error Sources

The performance of a strapdown inertial reference package is contingent upon the magnitude of errors from these principal sources:

(i) Initial alignment procedure.

(ii) Sensor measurements.

(iii) Computational algorithms.

The errors in the alignment procedure are associated with local, unknown anomalies and noise at the launch site, and measurement errors in optical and other launch site instrumentation, as well as the sensor errors.

The errors in the sensor measurements are functions of time, temperature, voltage, and environment. The sensor errors can be categorized as quasi-steady linear accelerations and angular rates of the vehicle. The gyro error model which has been used[4] includes non-g-sensitive drift (bias), g-sensitive drifts (mass unbalance), scale factor errors, and gyro input axis misalignments. The accelerometer error model includes bias, scale factor errors, and accelerometer input axis misalignment. The object of the calibration procedure is to determine the sensor error model coefficients in order to make possible in-flight compensation of the errors. A quasi-steady error budget for a boost-type mission must describe not only the basic accuracy of the sensors (or of the compensation) but also the time stability of the particular terms.

Once the sensor package has been mounted in the vehicle, most calibration coefficients cannot be found, since the strapdown calibration procedure requires the placing of the sensor package in a number of prescribed orientations relative to a reference rate vector and the local gravity vector. Good long-term stability of the sensor performance coefficients is essential.

A partial calibration of the strapdown system can be accomplished on the vehicle in conjunction with the alignment of the computational frame to the local earth coordinate frame of the launch point. Various procedures for mechanizing on-the-pad alignment of a strapdown inertial reference package can be used. These include pure optical alignment, accelerometer leveling plus optical azimuth alignment, and self-contained (gyrocompassing) alignment schemes. The terms that are calibrated are the biases of the two gyros (generally pitch and yaw) whose input axes at launch are nominally vertical, and the biases of the two level accelerometers. In the optically aided alignment schemes the gyro biases are automatically obtained, whereas in the self-contained alignment they are byproducts of the leveling loops. In both cases, the accelerometer biases can be obtained by filtering the system velocity outputs during a short period of navigation on the pad.

Table I lists the dynamic error sources, the equations, and the coefficients.

Evaluation of these dynamic error sources requires knowledge of the mission vibration environment and the specific sensor channel electronic loop characteristics. In the design of a strapdown inertial reference package, many of the requirements with regard to total system performance can be related directly to the requirements of the individual sensor loop dynamic response characteristics. A qualitative summary of the major factors which influence the overall design of a particular strapdown sensor loop configuration in light of mission requirements can be made as follows:

(i) The requirement on dynamic rectification errors due to vibratory inputs imposes a limitation on the compensation network and electrical gain at the frequencies of the expected vibrations. This implies that the closed loop float (pendulum) motions be kept to a minimum at these expected frequencies. In the ASA, the 1139B float motion is held to about 20 arc seconds; the pendulum in the Bell VIIB accelerometer moves only about 60 arc seconds.

(ii) The requirements for the tracking error, both steady-state and transient, imply a relatively high electrical gain to yield fast responding loops.

Besides the individual sensor loop dynamic errors, total system dynamic errors can exist due to the simultaneous processing of sensor outputs during vibrational inputs, and due to the noncoincidence of the sensor package with the center of mass of the vehicle. The functional forms of these type errors are also included in Table I.

The computational errors associated with a strapdown package present no basic limitation in navigation accuracy. Both the present and projected state of the art for digital computers permit repetition rates (of high order integration schemes) that exceed the bandwidth of the gyro sensors. This combination, in conjunction with sufficient word length, maintains the computational error below any required level. The other system computational error to be considered is that of sensor quantization which, in essence, places a time lag in the data stream (for angular data below the quantization level). Using time-modulated torquing, this pulse quantization can be made as fine as required, and this does not then constitute a significant system computational error.

The computation error types, round-off, quantization, and truncation, can be controlled by computer word length, pulse quantization magnitude, integration approach, and a practical cutoff point for high frequency inputs that are related to computer integration cycle time. The control of computational error is illustrated in Figure 7. Note that there is a trade-off available for any given set of system requirements between a possible shock mount design and the computer integration method and speed[5].

3. ACCEPTANCE TESTING

The production acceptance test program consists of operational, vibration, and thermal vacuum tests. During exposure to the test environments, performance parameters are monitored in each instrument channel.

The main requirements of the acceptance test are as follows:

(i) Verify the inertial performance characteristics of the SDIMU.

(ii) Evaluate the effects of vibration and thermal-vacuum environments by performing tests before, during, and after the environmental condition.

(iii) Observe the critical characteristics of the SDIMU in a sufficiently detailed manner so that the projected reliability and performance characteristics can be evaluated from a limited sample of data.

Operational testing is concerned primarily with calibrations. The parameters that must be determined are the scale factor, bias, input and spin axis, mass unbalance, and input axis alignment for each of the three gyro/PTSA channels, and scale factor, bias, and input axis alignment for each of the three accelerometer/PTSA channels.

The calibration procedure developed consisted of six static tests and six dynamic tests. In the static case, the package is held stationary, and data are gathered over a fixed time interval. The package is accurately aligned with respect to gravity and earth rate, permitting the evaluation of all performance parameters except gyro/PTSA scale factor and input axis alignments. Each input axis is aligned both vertically up and vertically down. The accelerometer/PTSA scale factor is obtained from the difference of the data accumulated with the input axis up and down, and the bias is determined from the sum of these data. The input axis alignment for each accelerometer/PTSA channel is obtained from the difference in outputs when the accelerometer input axis is horizontal. The gyro/PTSA bias is obtained from the sum of the data with the gyro oriented with its input axis up and down, and the spin axis mass unbalance for each gyro channel is determined from the difference in the data when each gyro is oriented with its spin axis vertically up and down.

Six dynamic tests are required for the evaluation of gyro/PTSA scale factor and the direction cosines. These tests consist of rotating the IMU about each gyro vertically oriented input axis at a constant rate over a precisely known angle, first in a clockwise and then in a counter-clockwise direction. For most applications, rates in excess of 1 degree/second must be sensed. Each gyro/PTSA scale factor is obtained from the difference in the data obtained from the clockwise and counter-clockwise directions. The input axis alignment is obtained from rotational data when the gyro is oriented with its input axis horizontal. A production test set for calibration is shown in Figure 8.

The vibration tests that are performed as part of the acceptance testing are designed to verify the functional operation of the SDIMU during a simulated mission level random vibration environment (see Figure 9). The SDIMU is attached to a vibration fixture in such a manner that the linear vibration input can be applied along any one of its primary axes. During vibration, critical performance parameters are monitored. Crystal accelerometers are positioned with their sensitive axes along the vibration input and located next to each SDIMU mounting point. Data from these instruments are recorded during the random vibration test to provide power spectral density spectra. The change in performance across the vibration test is determined both by observing the change in the mean value of the SDIMU calibration coefficients, obtained from calibrations performed before and after the test, and by observing the change in the "static" outputs of the sensor channels from immediately before to immediately after the test.

The objective of the thermal vacuum test is to verify the operation of the SDIMU when subjected to simulated mission temperature and vacuum environments. The SDIMU is installed on an adapter on a thermal shroud within a thermal vacuum chamber. After the chamber pressure is reduced to 1×10^{-5} torr or lower, the SDIMU is subjected to high (130°F) and low (30°F) temperature conditions for one hour at each test level. Two 200-second pulse counts are accumulated from the sensor channels every 15 minutes during and at the end of the one-hour period. The pulse counts accumulated at each time specified are compared against a pre-environment reference measurement in order to provide a measure of performance during the simulated mission environmental extremes. The change in performance across the thermal vacuum test is also determined by observing the change in the mean value of the calibration coefficients obtained from sets of calibrations performed before and after the test.

Acceptance test data for several production SDIMU's are presented in Tables II, III and IV. Table II presents a summary of SDIMU performance for mission-level vibration and thermal vacuum environments. These data show the total difference in the mean value of the calibration coefficients from before to after the environmental test, in terms of an average of the three channels per SDIMU. Improvement in gyro errors during thermal vacuum testing has been obtained recently by a change in the method of processing the surface of the gyro mounting seat to permit better temperature control.

Table III summarizes SDIMU stability data over the four-day acceptance test period. These stability data, expressed as 3σ values, *include* the effects of vibration and thermal vacuum testing. A technique has been developed to separate accelerometer bias from the bias of the rebalance amplifier. Analysis of the data shows that most of the bias shift is associated with the instrument, and that scale factor data for both gyros and accelerometers can be markedly improved by better control of noise in the power supply.

The short-term repeatability of SDIMU performance coefficients is defined as the root-mean-square of the standard deviations for each day's calibration test data. Six calibration tests are performed during each day.

Table IV summarizes the short-term data, which are principally of laboratory interest but are useful for comparative purposes.

4. SLED TESTING

4.1 Description of Test and Test Article

Recently a series of laboratory tests and rocket sled runs were conducted at Holloman Air Force Base, New Mexico. The test item was an early model of the LM/SDIMU with the accelerometer loops rescaled to a dynamic range of 18g without altering the bandwidth of the closed loop. In addition, an optical reference cube and mounting structure were attached to the package.

The objectives of the test program can be summarized as follows:

(i) Verify the ability of the LM strapdown system to maintain its functional integrity and to operate in the simulated missile environment of the AFMDC high speed test track.

(ii) Evaluate the operation of the strapdown inertial measurement unit and compare the output data with space/time reference data to determine the system accuracy.

(iii) Determine, if possible, error sources of the system with respect to gyros, accelerometers, and associated electronics.

(iv) Evaluate the computer algorithms planned for the system.

In all, a total of 14 sled tests divided into three phases (three Phase I runs, five Phase II runs, and six Phase III runs) were conducted. In addition, the SDIMU was laboratory tested and evaluated for static operational performance prior to sled tests to provide a base to which the sled test results could be compared.

The purpose of the Phase I runs was to establish a rotational and translational vibration profile. The objective of the Phase II runs was to check out and optimize the final system configuration for a qualitative performance evaluation. The Phase III runs were then performed with a detailed quantitative analysis of system performance as the objective.

The following is the general sequence of events in the Phase III sled test:

(i) Pre-run laboratory calibration of SDIMU and electronics.

(ii) Controlled transfer of SDIMU from laboratory to sled.

(iii) Pre-run optical alignment of ASA on sled.

(iv) Sled run and data gathering.

(v) Post-run optical alignment check of SDIMU on sled.

(vi) Controlled transfer of SDIMU from sled to laboratory.

(vii) Post-run laboratory calibration of SDIMU and electronics.

4.2 Environment

The Phase III sled test profile had a maximum linear acceleration of approximately 8g and a maximum velocity of 1500 ft/sec, occurring in the first 8 seconds of boost. Booster engine cutoff is followed by a decelerating coast of the sled toward a water brake approximately 27,000 ft down track. The water brake slows the sled velocity to approximately 200 ft/sec and the sled continues to coast to a stop at approximately 36,000 ft. A typical sled run profile is shown in Figure 10.

Power spectral density characteristics of the guidance pallet vibrations indicated that the most severe vibration levels occurred at the time of booster cutoff, with the highest level appearing as pitch vibrations of almost 10 degrees per second (RMS). Typical roll angular rate vibration spectra at various times during a sled run are shown in Figure 11. Linear acceleration vibration levels at booster cutoff are on the order of 1g (RMS).

4.3 Preliminary Results

The SDIMU sensor channel data were recorded on two test vehicle tape recorders during the sled runs. Simultaneously, signals from the track space/time reference displacement measuring system were recorded. The space/time system measures the times the sled passes the accurately surveyed light beam interrupters spaced along the 7-mile track. The recorded sensor data, the space/time reference, as well as all track survey data, are used to generate velocity error functions to indicate SDIMU performance.

Velocity error plots are usually obtained in either one of the following coordinate systems:

Track Coordinates - Velocity error functions are computed using the SDIMU accelerometer outputs only to obtain a quick-look indication of performance, by assuming that the pallet follows the track rails and has no high frequency oscillations due to sled angular motions and vibration isolator characteristics. Corrections are made for initial accelerometer alignment with respect to the track rails.

Astronomic Tangent Plane Coordinates - In this case, the ASA gyro channel data are used to transform the accelerometer channel data into astronomic tangent plane coordinates. The computed system velocity is compared to the space/time reference, resulting in error functions representing total SDIMU velocity error.

The data reduction used to obtain the velocity error functions utilized a CDC-3600 computer which was programmed to incorporate the strapdown algorithm. The system errors, therefore, include those of the computational scheme. Deterministic SDIMU sensor error coefficients of bias, scale factor, and input axis misalignment are compensated for in the computational scheme, using values obtained from pretest calibrations.

Preliminary results from the first of the Phase III sled runs are presented in Figures 12-14. The computed velocity error functions in track coordinates are shown in Figure 12. These error functions computed by subtracting the space/time reference from the compensated accelerometer outputs, exhibit two main effects: the curvature of the earth in the space/time reference along X (down track), and a tipping back (pitch rotation) of the ASA Z-axis (vertical) during boost. The space-time data are compensated to account for earth curvature, whereas the accelerometer outputs (in the absence of the gyro-dependent transformation function) are not. The tipping back of the Z-axis can be observed in the plot of Figure 13, which displays the angular displacement about the pitch axis as sensed by the ASA Y-gyro. The ASA appears to tip backward during boost, forward past its starting orientation (approximately 2 degrees elevation of X) during booster cutoff and coast, and then forward again upon entering the water brakes. Again, in the absence of the gyro-dependent coordinate transformation function, the tipping motions are not tracked by the system, and components of earth gravity and the sled thrust acceleration are erroneously reflected as velocity errors in the track coordinates.

The velocity errors computed in astronomic tangent plane coordinates in conjunction with the space/time reference are shown in Figure 14. Several significant error sources are displayed. The crosstrack (Y) velocity error is most sensitive to yaw (Z) gyro misalignment in tracking the direction of the thrust acceleration in the level plane. A misalignment of the yaw gyro allows the thrust velocity profile to be reflected in the crosstrack velocity error. The vertical (Z) velocity error is similarly sensitive to the pitch (Y) gyro performance in tracking the direction of the thrust acceleration in the vertical plane. The downtrack (X) velocity error is sensitive to the pitch gyro performance in resolving the directing of gravity.

It should be noted that the computational algorithms that were used to reduce the data of Figure 14 do *not* include gyro mass unbalance compensation. Since the output axis of the roll (X) gyro is perpendicular to the thrust acceleration, and since the crosstrack velocity error is also sensitive to the roll gyro drift, the magnitude of the X-gyro spin axis mass unbalance gains added importance. In addition, the output axes of the pitch (Y) and yaw (Z) gyros are perpendicular to gravity, thereby inducing a drift in pitch and yaw due to uncompensated Y-gyro input axis mass unbalance and Z-gyro spin axis mass unbalance, respectively.

Complete results of these tests will be described in a forthcoming Air Force technical report.

5. APPLICATIONS

The SDIMU which has been described in this paper lends itself admirably to spacecraft operations. Accelerometer bias can be updated in free flight if necessary; gyro drift is adequate for many existing missions. The relative ease of combining a strapdown package with an optical sensor to update bias adds to its usefulness. In an orbital transfer, the strapdown system shows some advantage over a platform. The vehicle roll axis is maintained perpendicular to the radius vector so that the sensor package will rotate at a constant rate with respect to inertial space. This condition serves to introduce a modulated drift rate (at an angular frequency equal to the pitch rate) which contributes to the attenuation of system burn velocity error due to sensor biases, and also reduces the divergence of the velocity error. This effect is illustrated in Figure 15, which displays the idealized functional behavior of system velocity error due to gyro bias for both a strapdown and a gimballed navigator.

In general, techniques for filtering and schemes for sensor error attenuation which are implemented in navigation systems using gimballed inertial measurement units may also be employed with the SDIMU system. Digital or pulse filters as processed in the flight computer will yield characteristics similar to those provided by analog filters employed in gimbal systems.

The preliminary sled test results indicate that, for boost and injection applications, the SDIMU is competitive with platforms. Pad alignment prior to launch is best handled optically for high performance, although gyrocompassing can be readily accomplished. Figure 16 shows injection errors as a function of coast angles based on 30-day sensor stability[7]. The ability to align the package in a preferred orientation with respect to the thrust vector tends to minimize gravity-sensitive errors. Platform errors tend to be compounded, since the angle between the thrust vector and the triad is changing. This advantage can only be realized, however, if the float motion in a strapdown gyro is held to very small values. Also, as previously noted the rebalance loops must have equal phase shift.

Considerable thought has gone into aircraft uses of an SDIMU. The high rates encountered in fighter aircraft produce an error which needs more engineering attention; however, there seem to be no inherent problems in minimizing this effect up to 200 degrees/second. For transport aircraft, inflight alignment of a Schuler-tuned system shows considerable promise. The use of electromagnetic position fixing makes the SDIMU very attractive. In addition, such a system makes possible the combining of both flight control and navigation sensory functions.

6. CONCLUSIONS

The following conclusions are drawn as a result of this strapdown system work which has been accomplished:

(i) High quality inertial system data, particularly suited for boost applications, can be obtained with a single-degree-of-freedom strapdown reference package.

(ii) The dynamics problem can be treated on a rigorous basis and, at this time, poses no barriers to further development.

(iii) Long-term stability and its relation to calibration and alignment is of paramount importance.

ACKNOWLEDGEMENTS

The author wishes to acknowledge the assistance of the Analysis Group of the Guidance and Control Department (under Mr J.Roantree) of United Aircraft Corporate Systems Center, and the cooperation of personnel at the Central Inertial Guidance Test Facility at Holloman Air Force Base. The work on the Abort Sensor Assembly for the Lunar Module was sponsored by NASA Manned Spacecraft Center at Houston and was subcontracted to United Aircraft through Grumman Aircraft Corporation and TRW Systems.

REFERENCES

1. *Scoville, A.E., Yamron, J. — *The Mechanization of a Strapdown Inertial System Based on Time-Modulated Torquing.* UACSC Report SCTM 248, August 1966.

2. Scoville, A.E. — *The Limitations in Maximum Gyro Torquing Rate.* UACSC Report SCTM 303, May 1967.

3. Lewellen, W.S. — *Analytic Solution for Fluid Stratification in a Gyroscope.* Journal of Space and Rockets, Vol.3, No.9, September 1966.

4. *Roantree, J., Kormanik, N. — *Generalized Design Criteria for Strapped Down Inertial Sensor Loops.* UACSC Report SCTM 246, August 1966.

5. *Sullivan, J.J. — *A Solution of the Critical Computational Problems Associated with Strapdown Navigation Systems.* UACSC Report SCTM 249, August 1966.

6. *Phase I Sled Test Report LM/ASA.* Air Force Missile Development Center, Interim Report No.1, AFMDC Technical Report MDC-TR-6759, March 1967.

---

* Presented at AIAA/JACC Guidance and Control Conference in Seattle, Washington, August 1966.

7. *Strapdown Inertial Navigator Integration Experiments.* United Aircraft Corporate Systems Center, UACSC Report SCTM 250, May 1966.

TABLE I

Summary of Dynamic Error Sources

●GYRO CHANNEL			
ERROR SOURCE	**ERROR EQUATION**	**COEFFICIENT**	**COEFFICIENT UNITS**
ANISOELASTIC	$\omega_D = \sigma \frac{A_{IA} A_{SA}}{2}$	$\sigma = \frac{M^2}{H}\left(\frac{1}{K_{SA}} - \frac{1}{K_{IA}}\right)$	$\frac{(deg/hr)}{(g)^2}$
IN-PHASE	$\omega_D = \sigma \frac{\Omega_{IA} \Omega_{SA}}{2}$	$\sigma = \frac{1}{H}(I_{SAF} - I_{IAF}) + \frac{1}{H} \lvert M \rvert_\omega (I_{SAW} - I_{IAW}) \cos\phi_\omega$	
ANISOINERTIA			$\frac{(deg/hr)}{(deg/sec)^2}$
OUT-OF-PHASE	$\omega_D = \sigma \frac{\Omega_{IA} \Omega_{SA}}{2}$	$\sigma = \frac{1}{H} \lvert M \rvert_\omega (I_{SAW} - I_{IAW}) \sin\phi_\omega$	
IN-PHASE	$\omega_D = \sigma \frac{\Omega_{IA} \Omega_{SA}}{2}$	$\sigma = \left\lvert \frac{\theta}{\Omega_{IA}} \right\rvert_\omega \cos\phi_\omega$	
SPIN-INPUT RECTIFICATION			$\frac{(deg/hr)}{(deg/sec)^2}$
OUT-OF-PHASE	$\omega_D = \sigma \frac{\Omega_{IA} \Omega_{SA}}{2}$	$\sigma = \left\lvert \frac{\theta}{\Omega_{IA}} \right\rvert_\omega \sin\phi_\omega$	
IN-PHASE	$\omega_D = \sigma \frac{\Omega_{SA} \Omega_{OA}}{2}$	$\sigma = \frac{J\omega}{H} \left\lvert \frac{\theta}{\Omega_{IA}} \right\rvert_\omega \sin\phi_\omega$	
SPIN-OUTPUT RECTIFICATION			$\frac{(deg/hr)}{(deg/sec)^2}$
OUT-OF-PHASE	$\omega_D = \sigma \frac{\Omega_{SA} \Omega_{OA}}{2}$	$\sigma = \frac{J\omega}{H} \left\lvert \frac{\theta}{\Omega_{IA}} \right\rvert_\omega \cos\phi_\omega$	
OUTPUT AXIS ANGULAR ACCELERATION	$\omega_D = \sigma \Omega_{OA}$	$\sigma = \frac{J}{H}$	$\frac{(deg/hr)}{(rad/sec^2)}$
SCALE FACTOR ASYMMETRY	$\omega_D = \sigma \Omega_{IA}$	$\sigma = \frac{S^{G+} - S^{G-}}{S^{G}_{r}}$	$\frac{(deg/hr)}{(deg/sec)}$
ROTOR UNBALANCE	$\omega_D = \sigma A_{SA}$	$\sigma = \frac{mr}{2H}$	$\frac{(deg/hr)}{(g)}$
FLOAT FRICTION SCALE FACTOR ERROR	$\omega_D = \sigma \Omega_{OA}$	$\sigma = \frac{2\mu r_B}{d\omega_m}$	$\frac{(ppm)}{(deg/sec)}$
●ACCELEROMETER CHANNEL			
ERROR SOURCE	**ERROR EQUATION**	**COEFFICIENT**	**COEFFICIENT UNITS**
IN-PHASE	$a_E = \sigma \frac{A_{IA} A_{PA}}{2}$	$\sigma = \left\lvert \frac{\theta}{A_{IA}} \right\rvert_\omega \cos\phi_\omega$	
VIBROPENDULOSITY			$\frac{\mu g}{g^2}$
OUT-OF-PHASE	$a_E = \sigma \frac{A_{IA} A_{PA}}{2}$	$\sigma = \left\lvert \frac{\theta}{A_{IA}} \right\rvert_\omega \sin\phi_\omega$	
ANISOINERTIA	$a_E = \sigma \frac{\Omega_{IA} \Omega_{PA}}{2}$	$\sigma = \frac{I_{PA} - I_{IA}}{P}$	$\frac{\mu g}{(deg/sec)^2}$
NON-LINEAR g^2	$a_E = \sigma \frac{A_{IA}^2}{2}$	σ = CONSTANT	$\frac{\mu g}{g^2}$
OUTPUT AXIS ANGULAR ACCELERATION	$a_E = \sigma \Omega_{OA}$	$\sigma = \frac{J_P}{P}$	$\frac{\mu g}{rad/sec^2}$
SCALE FACTOR ASYMMETRY	$a_E = \sigma A_{IA}$	$\sigma = \frac{S^{A+} - S^{A-}}{S^{A}_{r}}$	$\mu g/g$

Best Available Copy

TABLE I

Summary of Dynamic Error Sources (continued)

● SYSTEM			
ERROR SOURCE	**ERROR EQUATION**	**COEFFICIENT**	**COEFFICIENT UNITS**
NON-IDENTICAL GYRO LOOPS	$\bar{\omega}_{D_X} = \sigma \frac{\Omega_Y \Omega_Z}{2}$	$\sigma = \frac{1}{\omega}\left\lvert\frac{\Omega_{out}}{\Omega_{IA}}\right\rvert^2_\omega \sin(\phi_X - \phi_Y)$	$\frac{(deg/hr)}{(deg/sec)^2}$
PSEUDO-CONING			
OUTPUT AXIS ANGULAR MOTION	$\bar{\omega}_{D_X} = \sigma \frac{\Omega_Y^2}{2}$	$\sigma = \frac{J}{H}\left\lvert\frac{\Omega_{out}}{\Omega_{IA}}\right\rvert^2_\omega$	
PSEUDO-SCULLING	$a_{E_X} = \frac{\sigma A_Y \Omega_Z}{2}$	$\sigma = \sin\delta$	$\frac{\mu g}{(g)(deg/sec)}$
SIZE EFFECT	$a_E = \frac{\sigma \Omega^2}{2}$	$\sigma = X_o$	$\frac{\mu g}{(deg/sec)}$

ERROR MODEL NOMENCLATURE

IA, OA, SA	Input, output, spin axis
Ω	Angular rate amplitude
A	Linear acceleration amplitude
ω	Vibration angular frequency
$\bar{\omega}_D$	Average drift rate
M	Mass of the float-wheel combination
m	Effective mass unbalance of wheel
r	Effective mass unbalance displacement from wheel center of rotation
H	Spin angular momentum
J	Moment of inertia of the float about the output axis
I_{SAF}	Moment of inertia of the float about the spin axis
I_{IAF}	Moment of inertia of the float about the input axis
I_{SAW}	Moment of inertia of the wheel about the spin axis
I_{IAW}	Moment of inertia of the wheel about the input axis
I_{IA}	Combined moment of inertia of the float and wheel about the input axis
I_{SA}	Combined moment of inertia of the float and wheel about the spin axis
K_{SA}	Elastic compliance of float-wheel combination along the spin axis
K_{IA}	Elastic compliance of float-wheel combination along the input axis
$\left\lvert\frac{\Omega OUT}{\Omega IA}\right\rvert_\omega$	Gyro closed loop response evaluated at the input vibration amplitude and frequency
$\left\lvert\frac{\theta}{\Omega IA}\right\rvert_\omega$	Gyro closed loop float response evaluated at the input vibration amplitude and frequency
ϕ_ω	Closed loop phase shift evaluated at the input vibration frequency
$\lvert M_\omega \rvert$	Gyro motor transfer function magnitude evaluated at the vibration frequency along the spin axis
ψ_ω	Gyro motor transfer function phase evaluated at the vibration frequency along the spin axis
S^G	Average gyro scale factor
S^{G+}	Gyro scale factor for positive angular rate input
S^{G-}	Gyro scale factor for negative angular rate input
F	Gyro viscous damping
T_E	Fixed error torque about the output axis
$\hat{X}$	Unit reference axis of vehicle
r_B	Radius of the float bearings
d	Distance between the two float bearings
μ	Coefficient of static friction
ω_m	Maximum torquing rate of gyro
$\left\lvert\frac{\theta}{A_{IA}}\right\rvert_\omega$	Accelerometer closed loop float response evaluated at the input vibration amplitude and frequency
P	Accelerometer pendulosity
J_p	Moment of inertia of accelerometer pendulum about the output axis
S^A	Average accelerometer channel scale factor
a_E	Average accelerometer bias
δ	Phase difference between simultaneous linear and angular vibration
X_o	Displacement of accelerometer input axis from axis of rotation

Best Available Copy

TABLE II

SDIMU Performance for Mission-Level Environments

	VIBRATION TOTAL DIFFERENCE IN MEAN VALUE BEFORE TO AFTER VIBRATION. AVERAGE OF THREE CHANNELS PER SYSTEM				THERMAL VACUUM TOTAL DIFFERENCE IN MEAN VALUE BEFORE TO AFTER THERMAL-VAC. AVERAGE OF THREE CHANNELS PER SYSTEM			
SYSTEM S/N	CSC-2	005	006	007	CSC-2	005	006	007
GYRO BIAS (°/HR)	0.06	0.02	0.07	0.05	0.04	0.06	0.10	0.11
ACCEL. BIAS (μg)	19	9	12	16	14	14	25	16
GYRO SCALE FACTOR (PPM)	10	4	17	23	11	7	9	10
GYRO INPUT AXIS ALIGNMENT (ARC-SEC)	8	8	8	7	8	4	6	8
INPUT AXIS MASS UNBALANCE (°/HR/g)	0.06	0.09	0.02	0.02	0.05	0.14	0.02	0.07
SPIN AXIS MASS UNBALANCE (°/HR/g)	0.12	0.07	0.04	0.10	0.09	0.04	0.03	0.13
ACCEL. SCALE FACTOR (PPM)	5	6	7	6	4	2	3	18
ACCEL. INPUT AXIS ALIGNMENT (ARC-SEC)	8	6	8	2	7	3	4	8

TABLE III

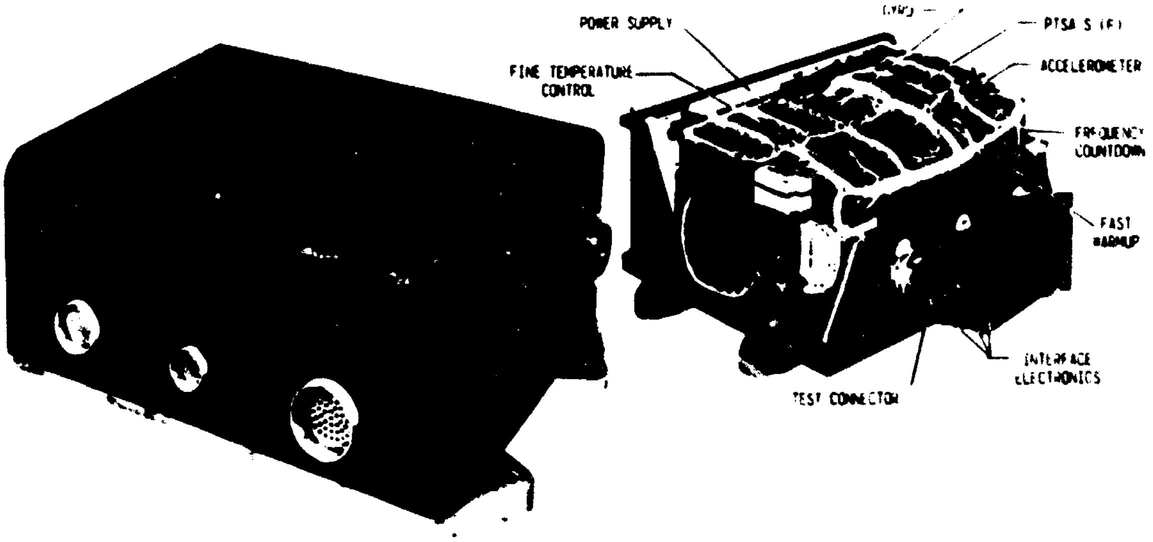
SDIMU 4-Day 3σ Acceptance Test Performance

	AVERAGE OF THREE CHANNELS				
ERROR SOURCES	CSC-2	CSC-3	005	006	007
GYRO CHANNEL					
Bias(°/hr)	0.18	0.10	0.17	0.36	0.25
Scale Factor (ppm)	38	51	22	60	68
IA Unbalance(°/hr)	0.19	0.11	0.32	0.09	0.12
SA Unbalance(°/hr)	0.32	0.21	0.32	0.18	0.17
IA Alignment(arc sec)	25	22	15	25	23
ACCEL CHANNEL					
Bias(μg's)	44	28	36	58	36
Scale Factor(ppm)	5	15	2	14	29
IA Alignment(arc sec)	12	26	12	24	16

TABLE IV

SDIMU Short Term Performance

PERFORMANCE PARAMETER	RMS OF FOUR DAILY 1 σ's
GYRO/PTSA BIAS (°/HR)	0.05
GYRO/PTSA SCALE FACTOR (PPM)	12.0
GYRO IA MASS UNBALANCE (°/HR/G)	0.01
GYRO SA MASS UNBALANCE (°/HR/G)	0.02
GYRO IA ALIGNMENT (ARC SEC)	3.0
ACCELEROMETER/PTSA BIAS (MICRO-G)	2.5
ACCELEROMETER/PTSA SCALE FACTOR (PPM)	2.0
ACCELEROMETER IA ALIGNMENT (ARC SEC)	2.0



EXTERNAL INTERNAL

Fig. 1 Abort Sensor Assembly

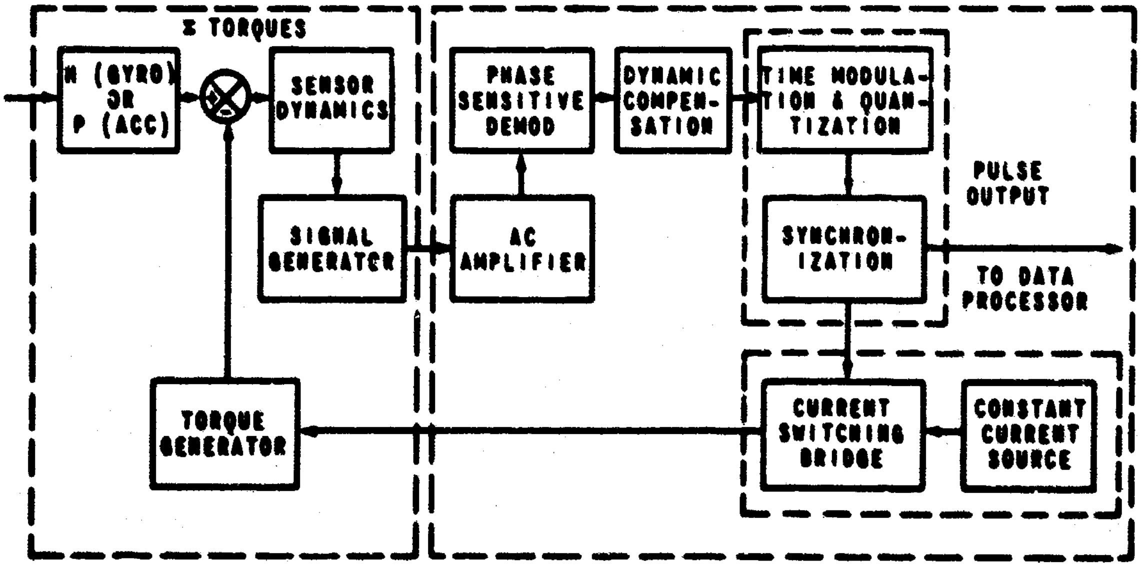


Fig. 2 Sensor loop functional diagram

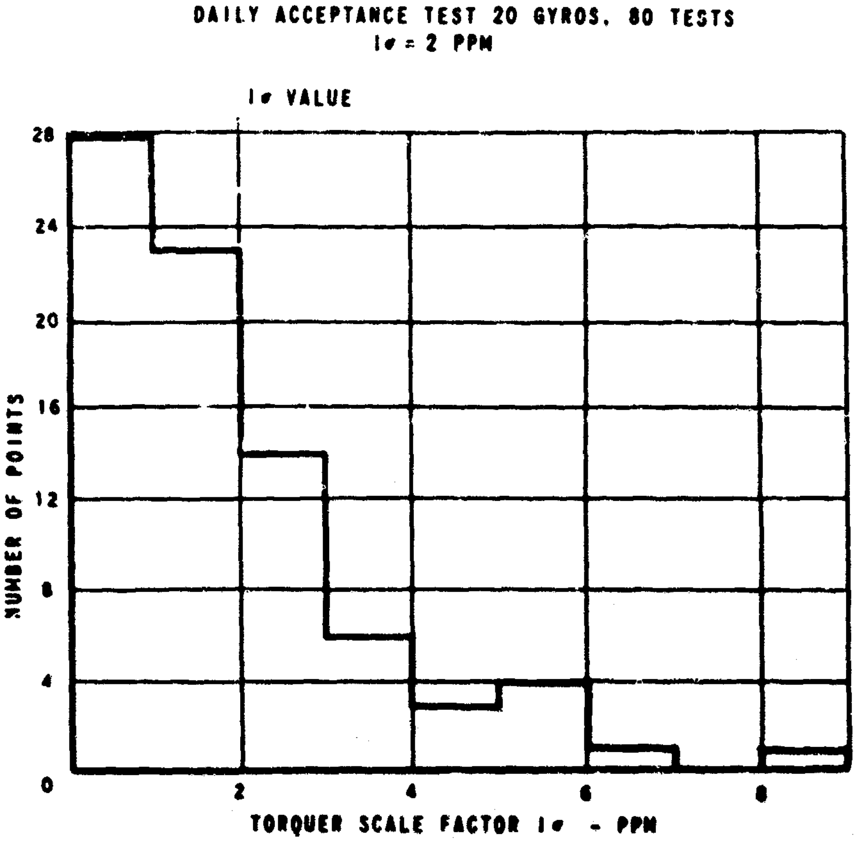


Fig.3 Torquer scale factory stability

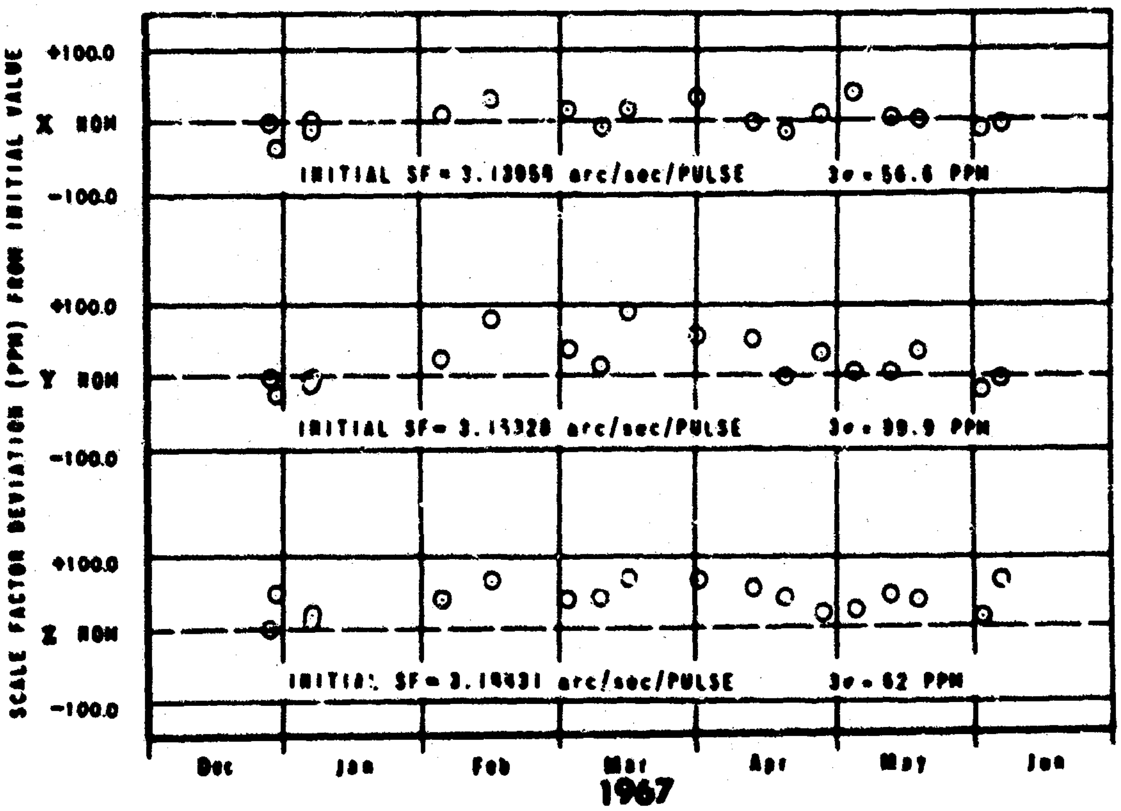


Fig.4 Torquer scale factor linearity

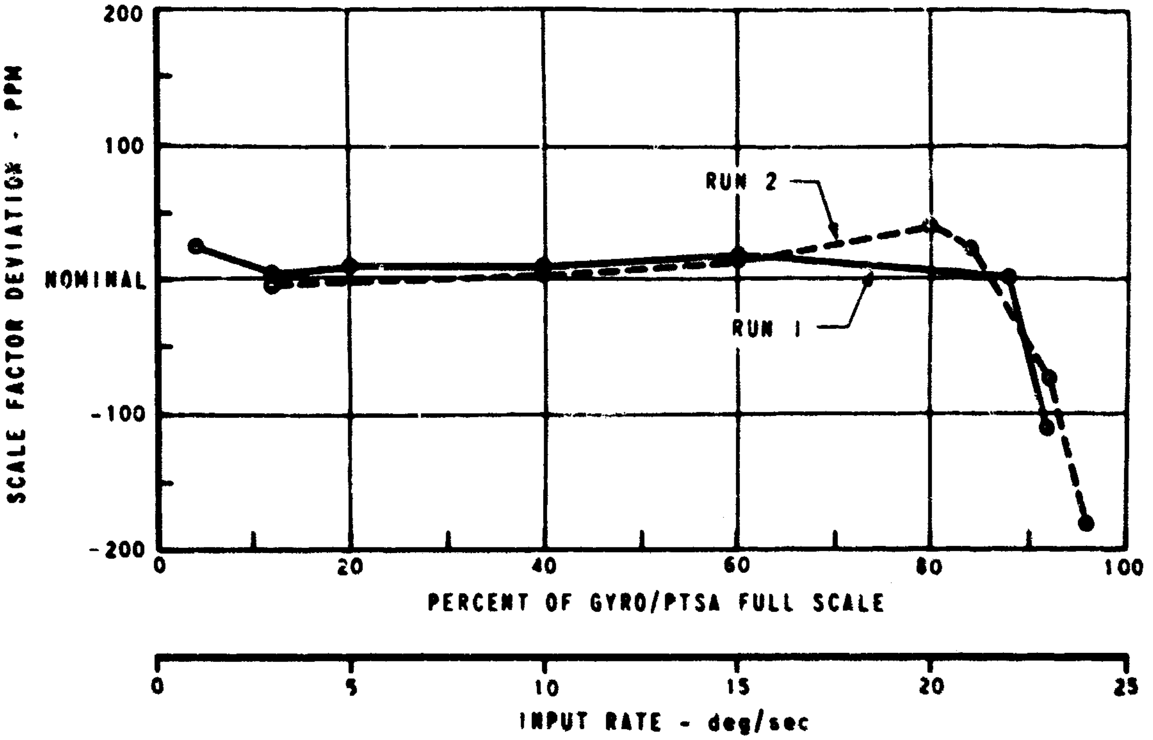


Fig.5 Gyro/PTSA scale factor

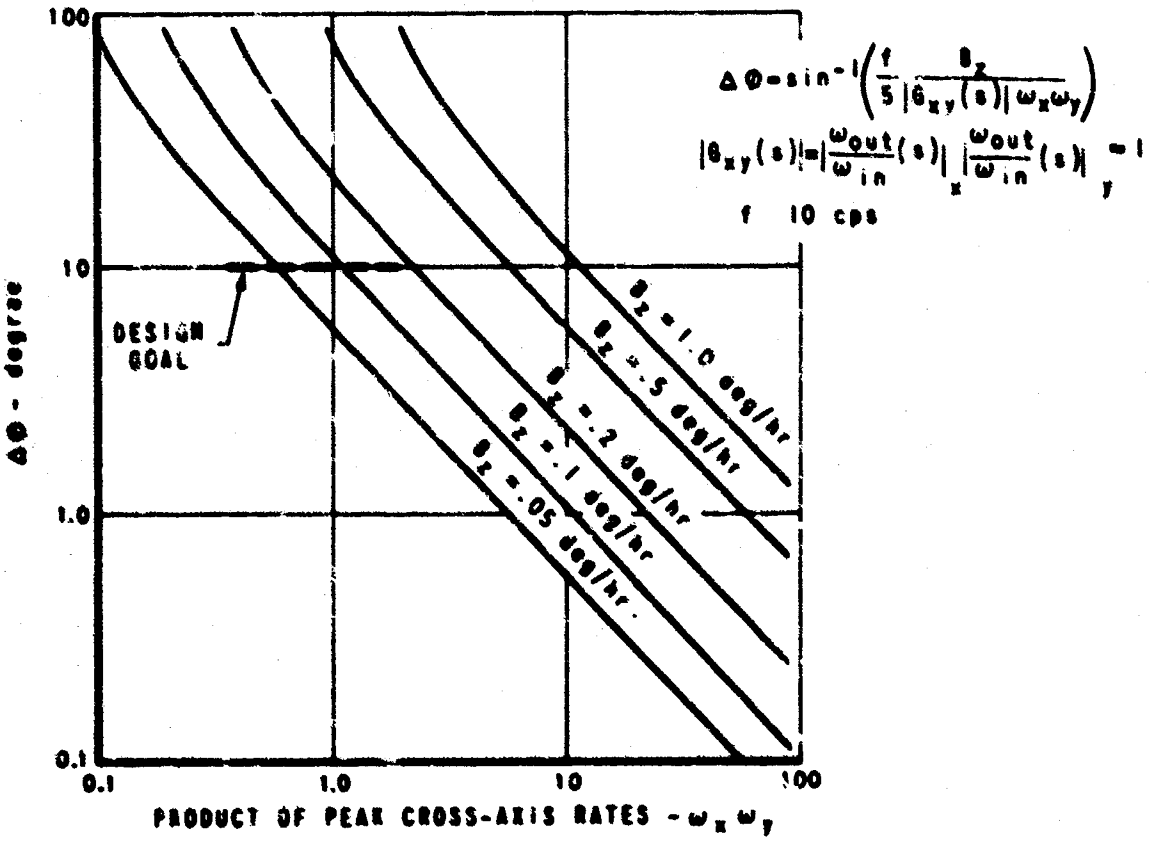


Fig.6 Gyro channel phasing matching

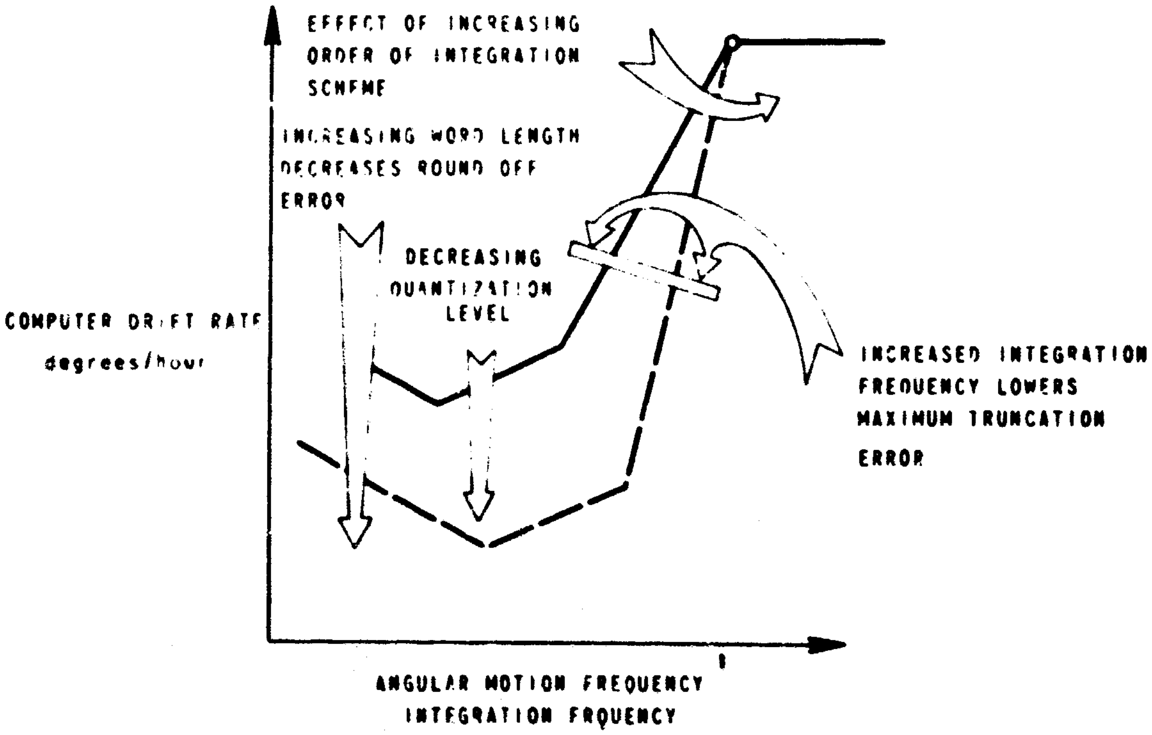


Fig.7 Control of computational error

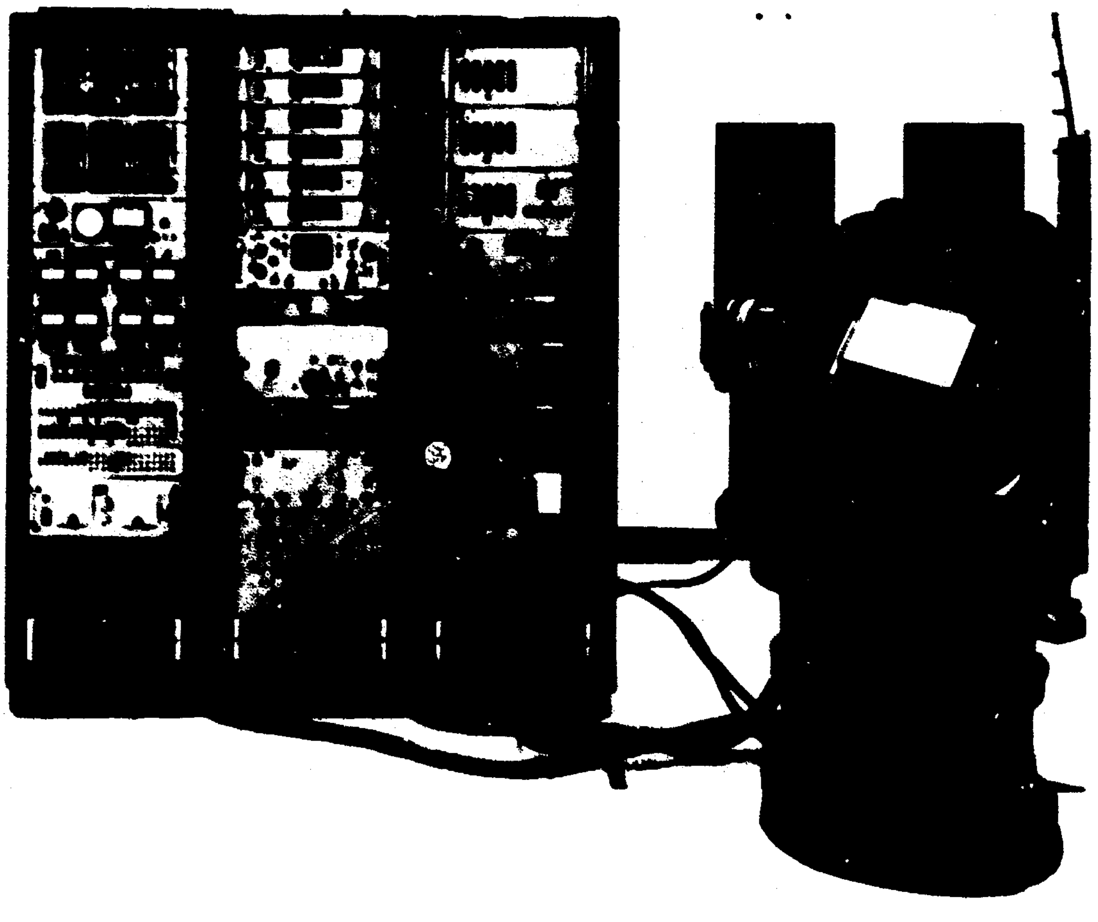


Fig.8 System test station

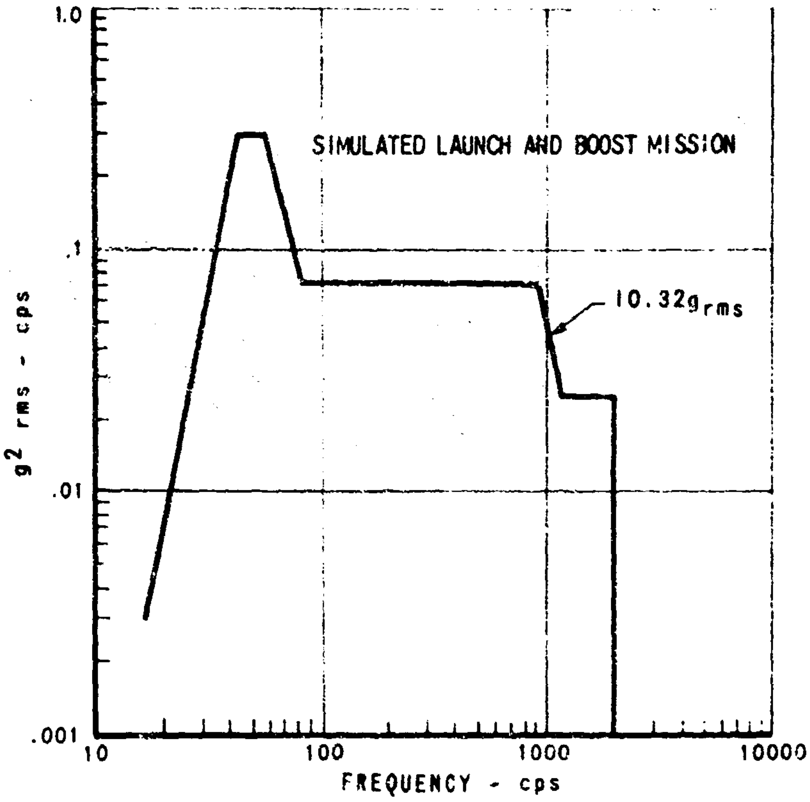


Fig.9 Vibration test profile

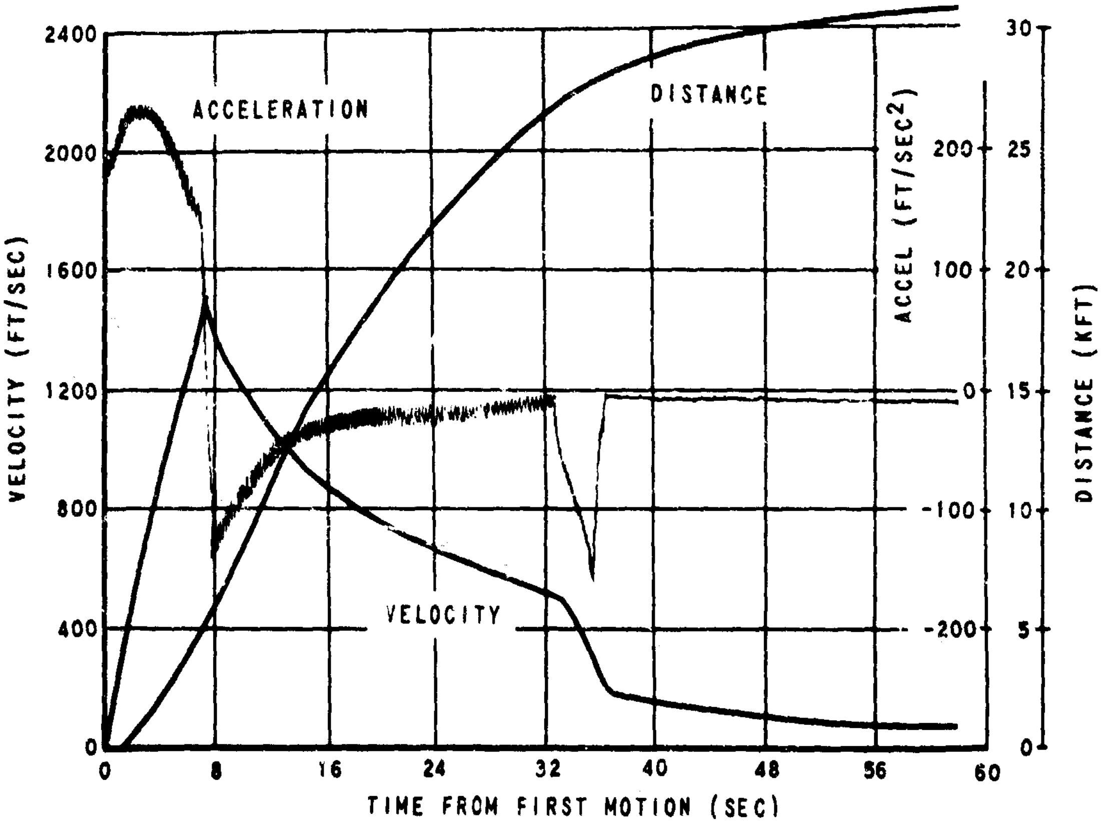


Fig.10 Sled run profile

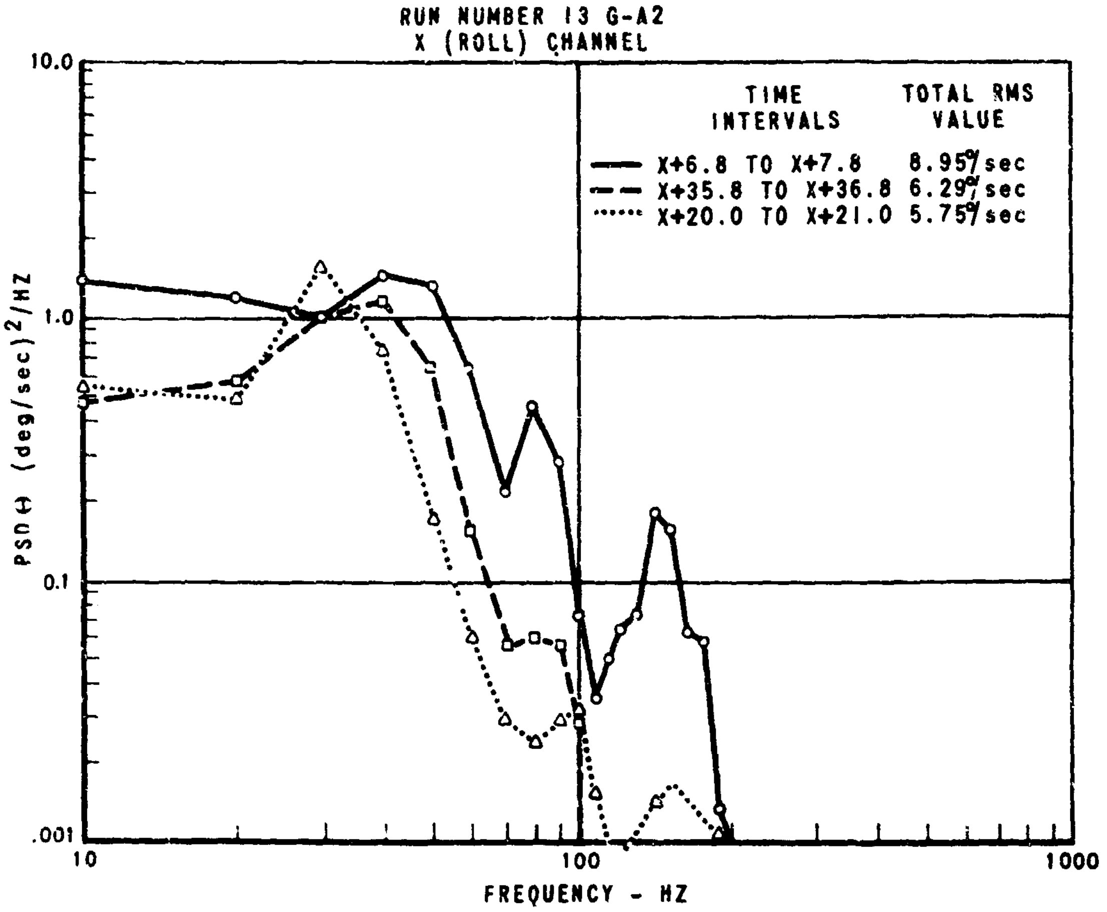


Fig.11 Sled angular power spectral density

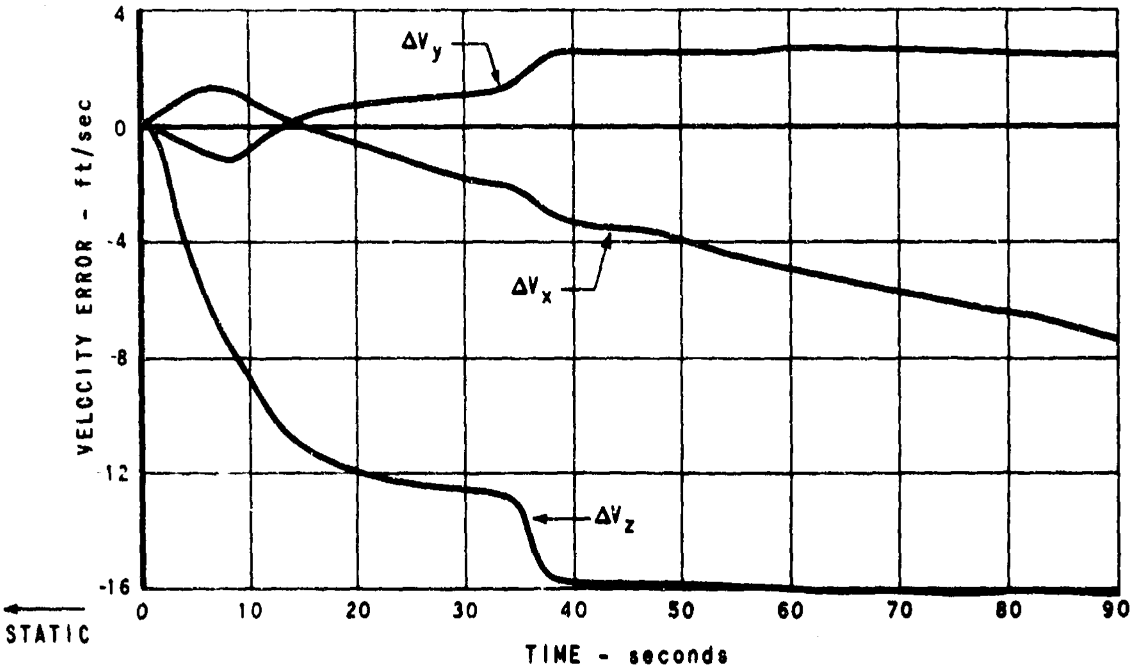


Fig.12 Sled velocity errors (accelerometer only)

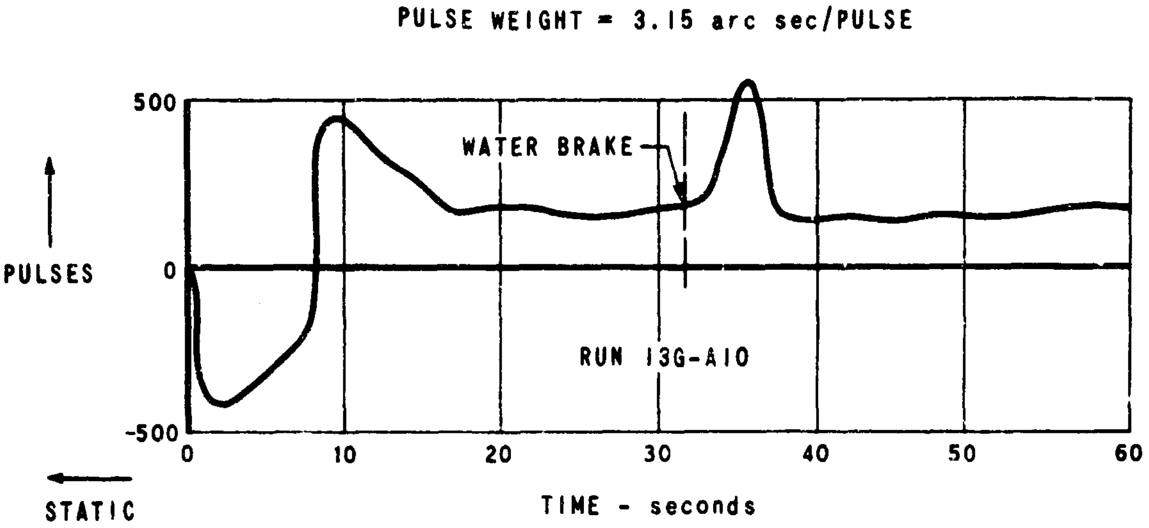


Fig. 13 Sled Y gyro pulse counts

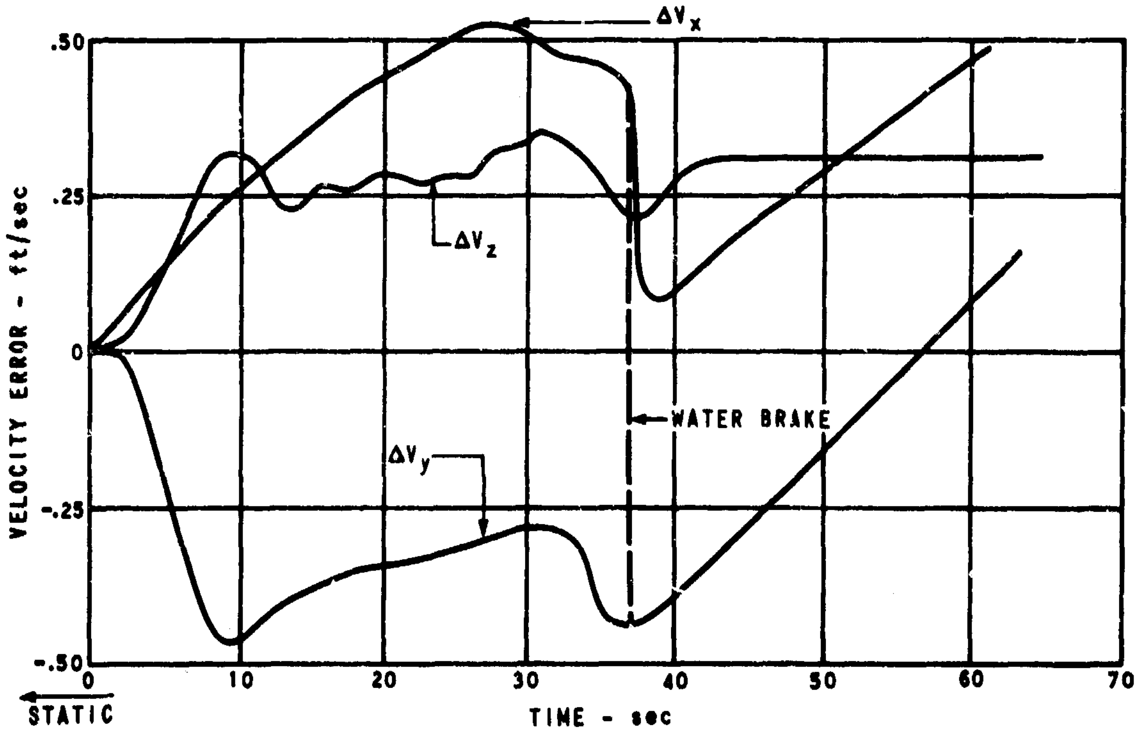


Fig. 14 Sled velocity errors (accelerometer and gyro)

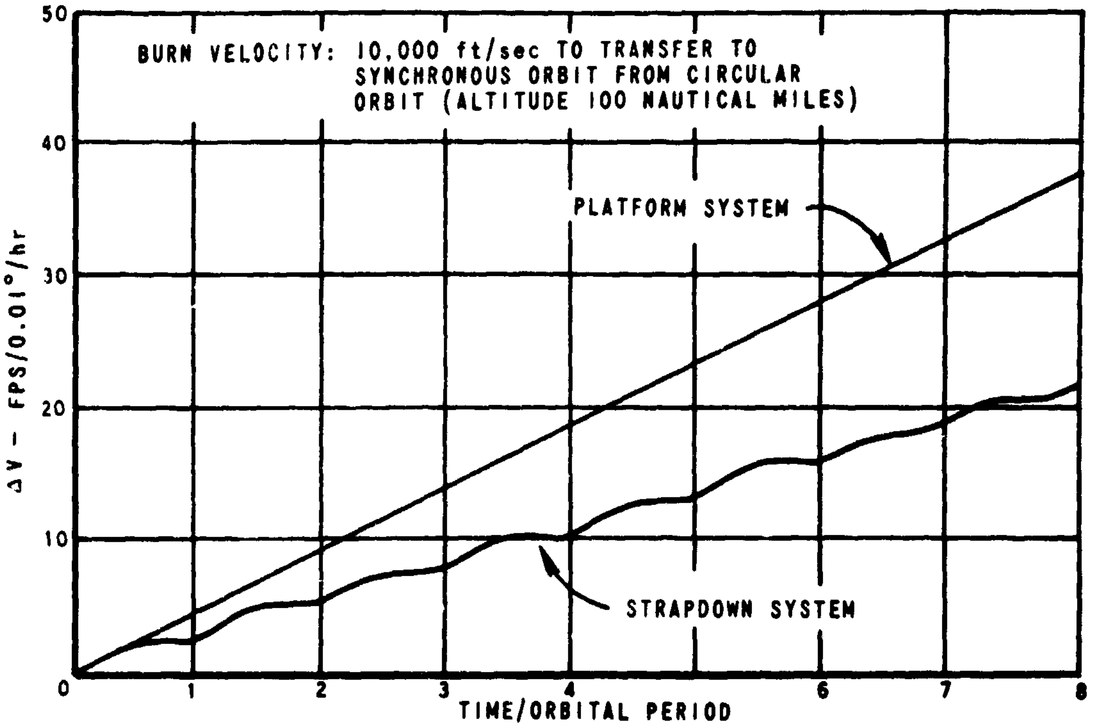


Fig. 15 Velocity error due to gyro bias

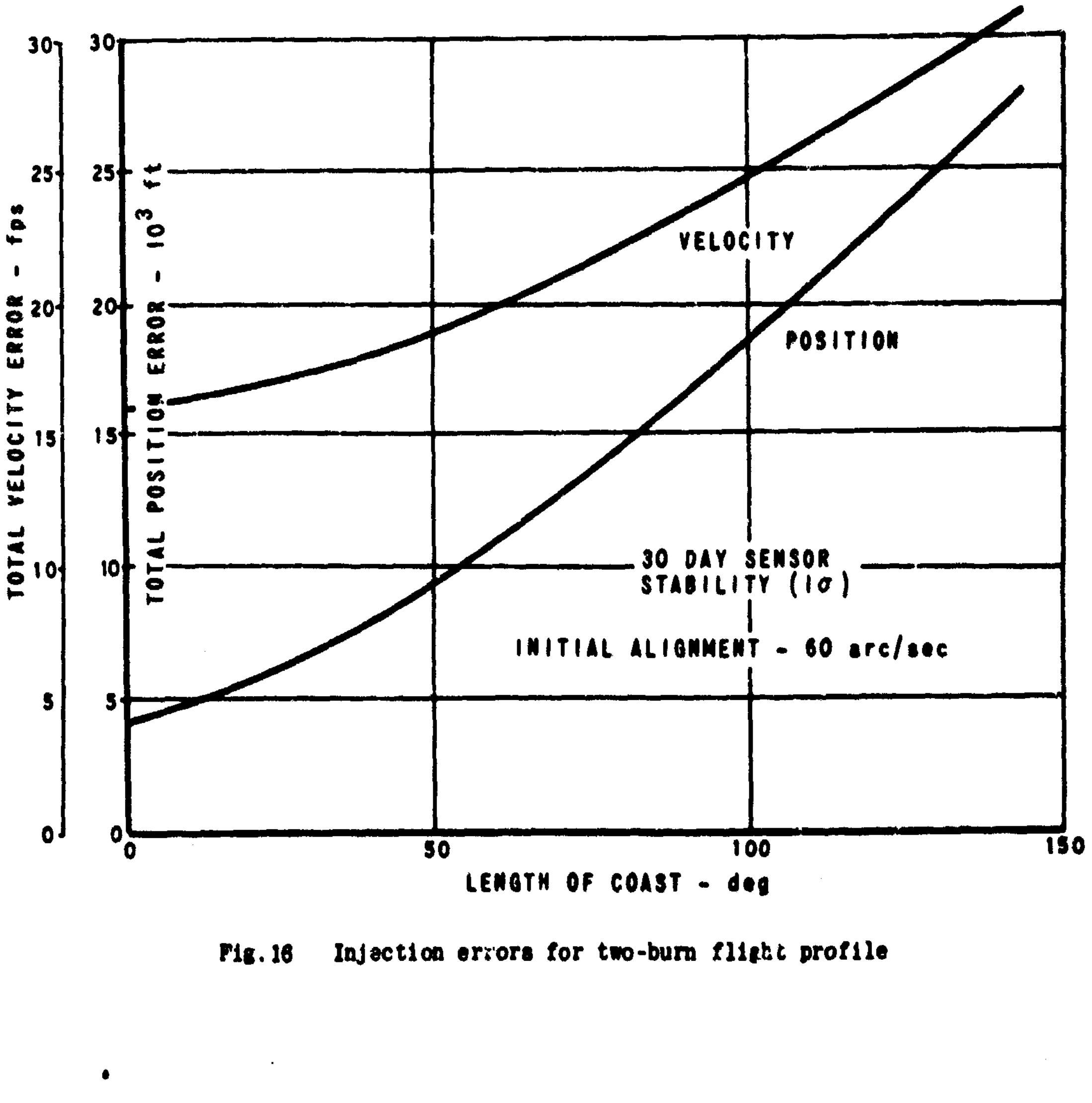


Fig. 16 Injection errors for two-burn flight profile

FIELD TEST, ALIGNMENT AND TRIMMING OF THE ELDO-A PLATFORM

by

Manfred Pütz

Fluggerätewerk Bodensee GmbH,
Überlingen, Bodensee, West Germany.

SUMMARY

The inertial platform used in the ELDO-A Program has to be tested and aligned prior to launch. This will be done by means of a Field Test and Alignment Unit, described in the paper. The Field Test Unit performs all the necessary switching between the platform system and the measuring and recording instruments. For easier operation and for the purpose of documentation all data are delivered digitally and printed. The applied method of the platform alignment starts with coarse alignment by synchro caging. Then the platform is leveled to the local vertical and subsequently aligned to North by closing the gyrocompassing loops. In this mode the calibration of the torquer scale-factors will be performed by aligning the platform in different positions. As a last step of the procedure the two vertical gyros will be trimmed, to achieve a higher alignment accuracy.

FIELD TEST, ALIGNMENT AND TRIMMING OF THE ELDO-A PLATFORM

Manfred Pütz

1. INTRODUCTION

In the ELDO-A Program, an inertial platform is used as a reference system for the attitude control of the vehicle. In a subsequent program this platform will be completed, as a full inertial navigation system, by the addition of a computer and the necessary interface equipment. In both cases, whether used for attitude reference only or as a part of an inertial navigation system, the platform has to be tested and aligned before launch. This can be done by means of a Field Test and Alignment Equipment, which allows a comprehensive testing of the platform, calibration of the torquer scale factors, trimming of the gyro drift and gyrocompassing alignment.

The platform testing is not part of the count-down. It will be undertaken, when the platform system is in the Preparation Building, to ensure that the system is working properly before it is installed in the vehicle. If a fault occurs, it can be localised to a certain extent by the test equipment, to determine whether the system can be repaired in the launching area or must be sent back to the manufacturer.

The alignment of the platform by gyrocompassing and the gyro drift-trim will also be done in the Preparation Building, at present. This alignment can be regarded as part of the test procedure, because the essential components of the platform, such as gyros and accelerometers, will be used for this purpose, and their accuracy can be checked. The actual alignment, of course, will be done when the platform is mounted in the vehicle.

2. PLATFORM TESTING

The Field Test Unit (Fig. 1) is rack-mounted in a console and consists of the following parts:

Function test unit
Recording and measuring unit
Three angle repeater unit.

In addition there is the alignment unit and the switch-on-unit; the latter, being delivered by the platform manufacturer, can also be built into the racks.

By means of the field test unit the following tests can be performed:

Insulation and continuity testing
Testing of platform electronics
Temperature measurements
Testing of gyro motors
Testing of gimbal servo motors and servo loops
Testing of gyros and accelerometers.

These tests will be done with the system operating at the nominal voltage, with undervoltage and overvoltage corresponding to the range given by the specifications.

The function test unit represents the central switching equipment, by means of which the complete field test is effected semi-automatically in relation to a specified program. After the cables of the platform, the electronics, and the switch-on-unit have been coupled to the test-equipment, the approximately 170 monitoring points of these systems are connected to the corresponding measuring and recording devices according to the sequence selected. This sequence is given by the test program. The appropriate connections are provided automatically by operating push-buttons. In addition, the operating voltage necessary for the tests is switched to the system to be tested.

The measured values of the resistances, voltages, currents, and frequencies are delivered digitally. For the purpose of documentation, and for proper identification of each measurement, the test point number, measured value, and dimension are printed out. The measured transient responses are recorded. These are the transient responses of the gimbal servo loops and the accelerometer servo loops, the output signal of the accelerometers when rotating the cluster about the different axes, and the platform temperature.

Switching is done in the relay-unit, which contains 2 relay matrices of 200 Hercon relays, 2 diode matrices, and an electrical dummy of the platform. This dummy is used to test the platform electronics before connecting it to the platform. By this means platform damage due to faults in the platform electronics is avoided. Whereas the platform electronics can be repaired in the launching area at Woomera, the platform has to be sent back to the manufacturer in Europe if a fault occurs.

In the relay-unit of the Field Test Equipment, as well as the 200 Hercon relays, there are additional relays belonging to safety circuits which will protect the platform system in the event of incorrect operation. The 2 diode matrices deliver the test number and dimension of the measured value to the printer in BCD code.

As experience during the operations in Woomera has shown, the Field Test Equipment allows comprehensive testing of the platform system in a relatively short time.

3. GYROCOMPASSING ALIGNMENT

As already mentioned, the alignment of the platform before launch is done by gyrocompassing. The platform used in the ELDO-A Program for attitude reference is a four gimbal SF 600 made by Ferranti Ltd., of Edinburgh. The relations given in Table I exist between the axes of the platform, the axes of the vehicle, and those of the earth-fixed coordinate system in the position of the initial vehicle, the platform axes being numbered from inside to outside.

TABLE I

Platform axis	*Vehicle axis*	*Earth-fixed coordinate system*
No. 1 (innermost axis)	Pitch	E-W
No. 2	Redundant roll axis	Vertical
No. 3 (intermediate axis)	Yaw	N-S
No. 4 (outer axis)	Roll	Vertical

The platform contains 3 single-axis floated gyros. Type SAGEM 10 675 A, and 3 single-axis accelerometers, Type SAGEM 10 625 A. As the natural frequency of the servo loops is several orders of magnitude higher than that of the alignment loops and the stiffness is very high, the function of the servo loops can be neglected for the present, i.e. it is assumed that the platform is exactly aligned at any time with respect to the gyro axes. The inertial components have the accuracy given in Table II.

TABLE II

Gyro SAGEM 10 675 A	
Mass unbalance drift	1^{o}/h/g (max)
Fixed restraint drift	1^{o}/h (max)
Anisoelastic drift	0.04^{o}/h/g^{2} (max)
Random drift O.A. vertical	0.03^{o}/h (1σ)
Random drift O.A. horizontal	0.05^{o}/h (1σ)
Torquer scale factor	2^{o}/h/mA^{2}
Accelerometer SAGEM 10 625 A	
Range	± 20g
Linearity	10^{-4} for ± 1g
Bias	2×10^{-4} g
Bias discrepancy	0.4×10^{-4}g
Zero stability (day to day)	2×10^{-4}g

The configuration of all the elements for the essential alignment is illustrated in Figure 2, where the 3 gyros are shown with their relationship of axes and the 2 horizontal accelerometers are represented by pendulums. The x and y axes of the coordinate system used are horizontal, the y axis points to the North, and the z axis is vertical. The earth's rotation ω_e is resolved into its two components along the y and z axes, $\omega_e \cos\lambda$ and $\omega_e \sin\lambda$. A further resolution of the $\omega_e \cos\lambda$ component relative to the axes of a coordinate system rotated about the z axis by an angle ϑ_z yields the components $\omega_e \cos\lambda \cos\vartheta_z$ (along the input axis of the NS gyro) and $\omega_e \cos\lambda \sin\vartheta_z$ (along the input axis of the EW gyro).

Vertical alignment of the platform is achieved by connecting the outputs of the EW and/or the NS accelerometer across filters and amplifiers with the torquers of the NS and/or EW gyros. Two simple control loops are thus created, in which the effect of the disturbances must be kept sufficiently low by means of suitable filters and proper selection of the parameters.

The horizontal component of the earth's rotation, $\omega_e \cos\lambda \cos\vartheta_z$, represents the predominant disturbance factor, and should therefore be compensated. As in the aligned condition ($\vartheta_z = 0$) its value is constant and varies only slightly in the vicinity of this point ($\omega_e \cos\lambda \cos\vartheta_z$), such a compensation can be effected by a constant bias signal on the torquer of the NS gyro. This requires, however, that the scale factor of the torquer is known and constant. A further possibility of compensating this disturbance factor is the use of an additional integrator in the loop. This eliminates not only the effect of the earth's rotation but also that of the constant amount of gyro drift on the accuracy of the vertical alignment. A disadvantage of using an integrator is the higher sensitivity to accelerations caused by vehicle motion.

To achieve alignment to North, a connection between the NS accelerometer and the azimuth gyro is made, as before, by means of a suitable electronic unit. If there is a deviation of the cluster-fixed coordinate system ξ, η, z, from the earth-fixed coordinate system x,y,z by the angle ϑ_z , the component $\omega_e \cos \lambda \sin \vartheta_z$ of the earth's rotation appears as a disturbance factor for the vertical alignment loop, monitored by the NS accelerometer. This results in a vertical deviation proportional to $\omega_e \cos \lambda \sin \vartheta_z$. A corresponding signal is thus applied to the azimuth gyro which disappears when $\vartheta_z = 0$ and the platform is thus aligned to North. Inverting the sign in the loop results in a state at $\vartheta_z = 180^0$, which corresponds to an alignment of the platform to South. The vertical component of the earth's rotation represents a disturbance factor for the azimuth loop, which can again be either compensated by a constant bias signal or made ineffective by using an integrator in the azimuth loop. Figure 3 shows the root loci of a system with first-order filter with and without integrator. It can be clearly seen that, using an integrator, a certain damping level of the system cannot be exceeded by reducing the gain factor R_0 , as can be done in the system without integrator.

The block diagrams of the gyrocompassing loops (Fig.4) can be derived from the system equations. They show that the mode of operation of the system can be represented by an independent vertical alignment loop and a coupled vertical and azimuth alignment loop. From the 3 components of the earth's rotation the first two components, $\omega_e \cos \lambda \cos \vartheta_z$ and $\omega_e \sin \lambda$, are disturbance factors, whereas the component $\omega_e \cos \lambda \sin \vartheta_z$ is the control input producing alignment of the system to North.

4. ALIGNMENT AND TRIMMING PROCEDURE

The gyros used in the platform SP 600 were originally developed for an application not requiring an accurate torquer. For this reason, both the linearity and the long-term stability are not very good with respect to the gyrocompassing requirements. In order to obtain maximum accuracy from the components, the gyro torquer will be calibrated immediately before the actual alignment and trimming procedure.

The fundamental idea is as follows. When the platform lies within the alignment loop and a steady state has been attained, i.e. when the alignment operation is completed, a current corresponding to the horizontal component of the earth's rotation, $\omega_e \cos \lambda \cos \vartheta_z$, must always be fed into the NS gyro torquer. As the value $\omega_e \cos \lambda \cos \vartheta_z$ for $\vartheta_z = 0$ is known exactly for the given latitude, the scale factor representing the proportionality constant between the current and the resulting rotation could be determined by measuring the current in the torquer.

The special problem is now that the drift, which is still unknown, also acts upon the gyro, in addition to the earth's rotation. Two measurements are therefore necessary to separate these two quantities. For the stationary case, the current in the torquer is

$$K_T I_{T_1} = \omega_e \cos \lambda \cos \vartheta_z + \omega_D + (\vartheta_{z0} + \epsilon_{NS})\, \omega_e \sin \lambda$$

when the system is aligned to North. The term $(\vartheta_{z0} + \epsilon_{NS})\omega_e \sin \lambda$ results from coupling induced by the vertical alignment error ϑ_{z0} and the misalignment between the inertial components ϵ_{NS} .

If the system is then aligned to South, the sign of the horizontal component of the earth's rotation has changed. Provided that the drift remains constant during rotation of the platform, the current in the torquer is

$$K_T I_{T_2} = -\omega_e \cos \lambda \cos \vartheta_z + \omega_D + (\vartheta_{z0} + \epsilon_{NS})\, \omega_e \sin \lambda \, .$$

The drift and the coupling errors can therefore be eliminated from these two equations to determine the scale factor.

The accuracy limits of this measurement depend on the value of the random drift. Moreover, errors can occur when the trimming state of the gyros changes during rotation from North to South. This may occur particularly when summing networks are used. The gyro should be pre-trimmed before measurement, in order to avoid excessive deviation errors from North.

For small angles ϑ_z , the term $\cos \vartheta_z$ can be neglected. The error thus caused is approximately $1°/oo$ for $\vartheta_z = 2.5°$.

For scale factor calibration of both vertical gyros, the platform has to be aligned successively in the positions $\vartheta_z = 0°$, $90°$, $180°$ and $270°$, and the current flowing through the torquer of the gyro whose input axis is actually positioned in the NS direction, must be determined. Elimination of the still unknown gyro drift is possible from the four measured values, using the given relations, permitting the calculation of the two torquer scale factors.

The next step is the trimming of the two vertical gyros. The gyro drift consists of different components, which are acceleration-independent, acceleration-dependent and random. For the NS gyro and the EW gyro, which are positioned with their output axes vertical, the acceleration-dependent drift is zero and only the two other components are present.

A great advantage of gyrocompassing alignment is that the constant amount of the drift, which may change from switch-on to switch-on, can be trimmed before the actual application of the platform. This trimming is done during the alignment of the gyro, whose input axis is positioned in the NS direction. For this purpose, the equations for the torquer current in the steady state are again used:

$$K_T I_T = \omega_e \cos \lambda \cos \vartheta_z + \omega_D + (\vartheta_{zc} + \epsilon_{NS})\, \omega_e \sin \lambda ,$$

i.e., the current generates a gyro precession corresponding the sum of the horizontal component of the earth's rotation, the drift and the fraction of the vertical component of the earth's rotation induced by coupling errors. The gyro trimming can now be done so that the appropriate trim potentiometer of the NS gyro is adjusted until the torquer current of the NS gyro is

$$I_T = \frac{1}{K_T}(\omega_e \cos \lambda \cos \vartheta_z + (\vartheta_{ze} + \epsilon_{NS})\, \omega_e \sin \lambda)$$

and the active gyro drift $\omega_D = 0$. The trimming will be done by neglecting the terms $\cos \vartheta_z$ and $(\vartheta_{ze} + \epsilon_{NS})\, \omega_e \sin \lambda$. The resulting errors in drift trimming are, for $\lambda \approx 45°$ and

$$\vartheta_z = 2.5° \text{ (EW gyro only pre-trimmed)}, \quad \Delta\omega_D \approx 0.01°/h$$

$$(\vartheta_{ze} + \epsilon_{NS}) = 1 \text{ mrad} , \quad \Delta\omega_D \approx 0.01°/h$$

For trimming the second vertical gyro, the cluster has to be rotated through $90°$.

5. EFFECT OF THE COMPONENT ERRORS

The different component errors can affect both the dynamic response and the static accuracy of the system.

The first group includes the errors due to misalignment of the inertial components on the platform. They lead to couplings between the individual loops, as well as to additional cross-coupling of the variables within these loops. Figure 5 illustrates all the resulting couplings of this kind.

The following definitions of misalignment errors are used:

ϵ_{NS} NS gyro and NS accelerometer

ϵ_{EW} EW gyro and EW accelerometer

ϵ_{AZ} AZ gyro and AZ accelerometer

δ_{NS} NS gyro and AZ axis of platform

δ_{EW} EW gyro and AZ axis of platform.

A closer examination of these effects shows, however, that a misalignment between the input axis of the EW gyro and the AZ axis of the platform has the greatest effect on the dynamic response of the system. This causes the EW-gyro to sense a component of the cluster rotation about the azimuth axis which results in a component additional to the control signal $\omega_e \cos \lambda \sin \Phi_z$ as a function of the speed $\dot{\Phi}_z$. In the steady state we have $\dot{\Phi}_z = 0$, i.e., the static accuracy of the system is not affected. Because of this effect between the output and input value of the control loop there is a proportionally acting element $H\omega_e \cos \lambda$, and parallel to it a damping element with the transfer function $H\delta_{EW}s$. Depending on whether the alignment is made to North (0^0 position) or to South (180^0 position), the sign of this element changes, whereas that of the proportional element remains unchanged and the damping will be higher or lower compared to a system without coupling.

It should be noted that similar effects can also be caused by electrical couplings of circuits.

The dominant quantities on which the accuracy of the system depends in the steady state are

> gyro drift, the earth's rotation, zero stability of electronics, and accelerometer bias.

The resulting errors are listed in Table III

TABLE III

Vertical Alignment	
Accelerometer bias 2×10^{-4}g	40″
Amplifier drift 0.3 μV/°C	3×10^{-2}″/°C
Earth rotation $\omega_e \cos \lambda$	----
Constant amount of gyro drift (EW and NS)	----
Azimuth Alignment	
Accelerometer bias 2×10^{-4}g	1′
Amplifier drift 0.3 μV/°C	----
Random drift of EW gyro 0.03°/h	10′
Torquer linearity of EW gyro 0.5%	15′
Constant amount of gyro drift (AZ)	----
Earth rotation $\omega_e \sin \lambda$	----
Misalignment of inertial components 1 mrad	3′

It can be seen from Table III that the vertical alignment accuracy mainly depends on the accelerometer bias. The accuracy of the azimuth alignment is determined by two dominant error sources, the random drift and the torquer linearity of the EW gyro.

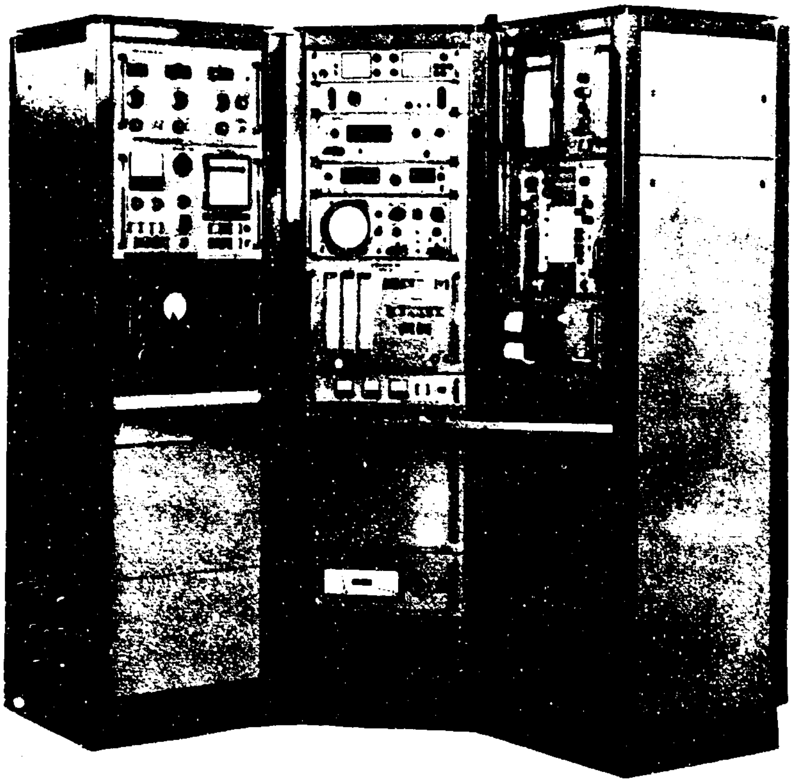


Fig.1 Platform check-out and alignment equipment

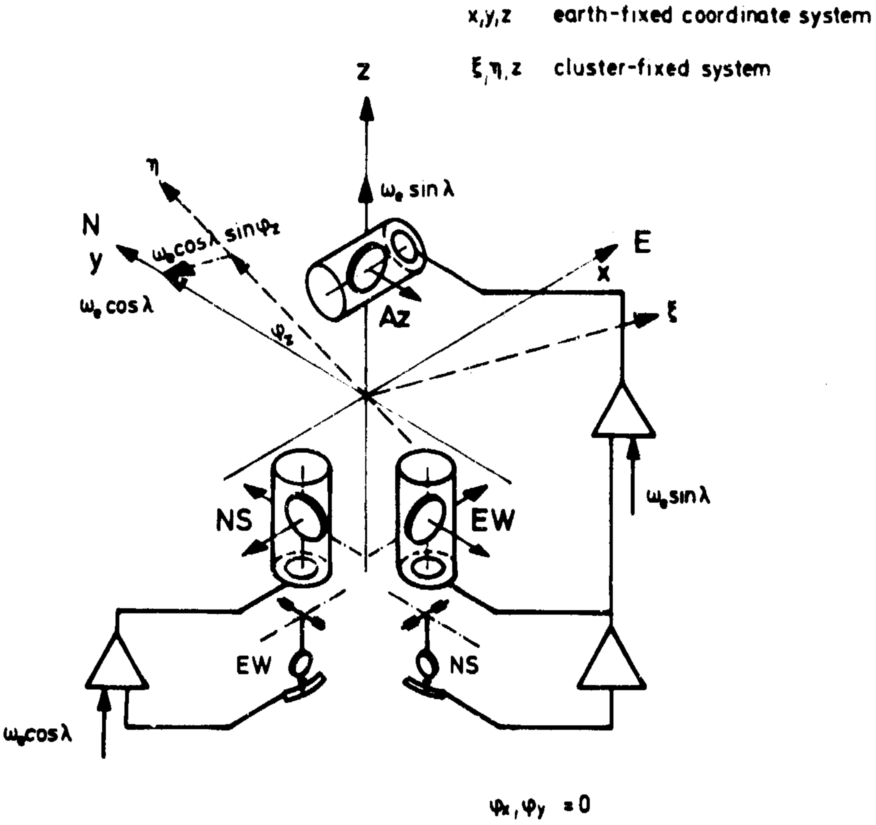


Fig.2 Simplified gyrocompassing system

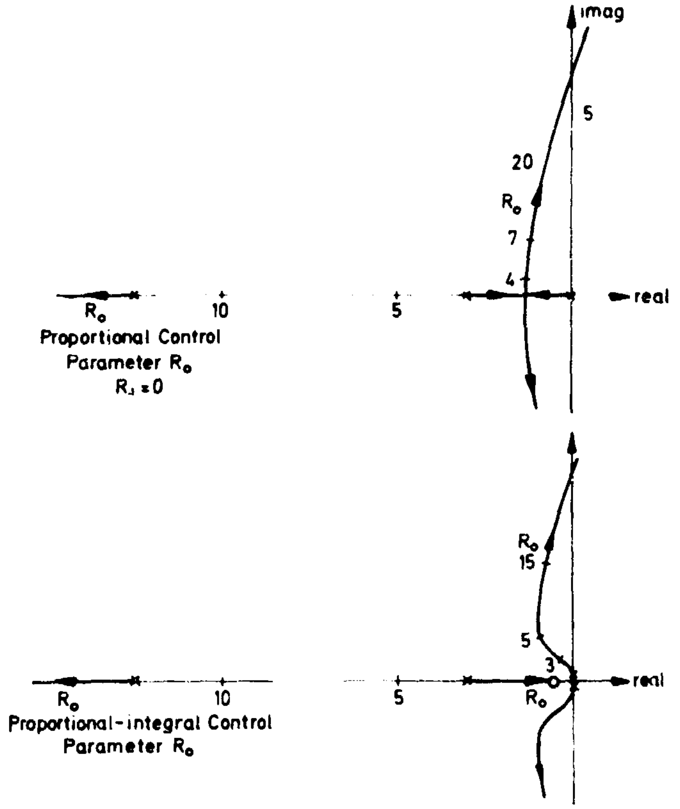


Fig.3 Root locus curves of gyrocompassing system

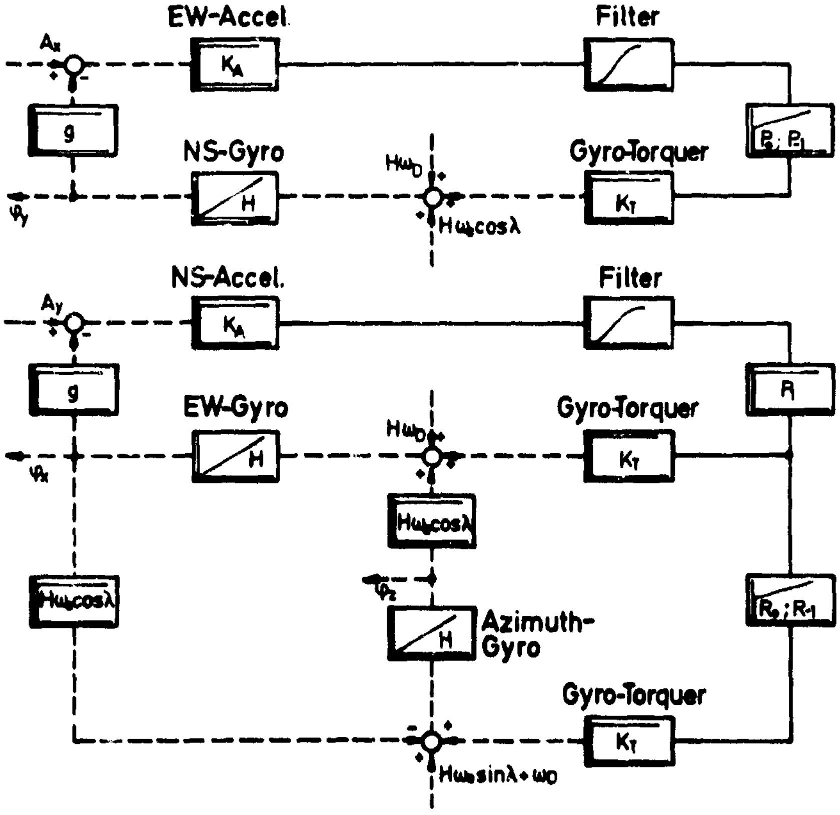


Fig.4 Block diagrams of alignment loops

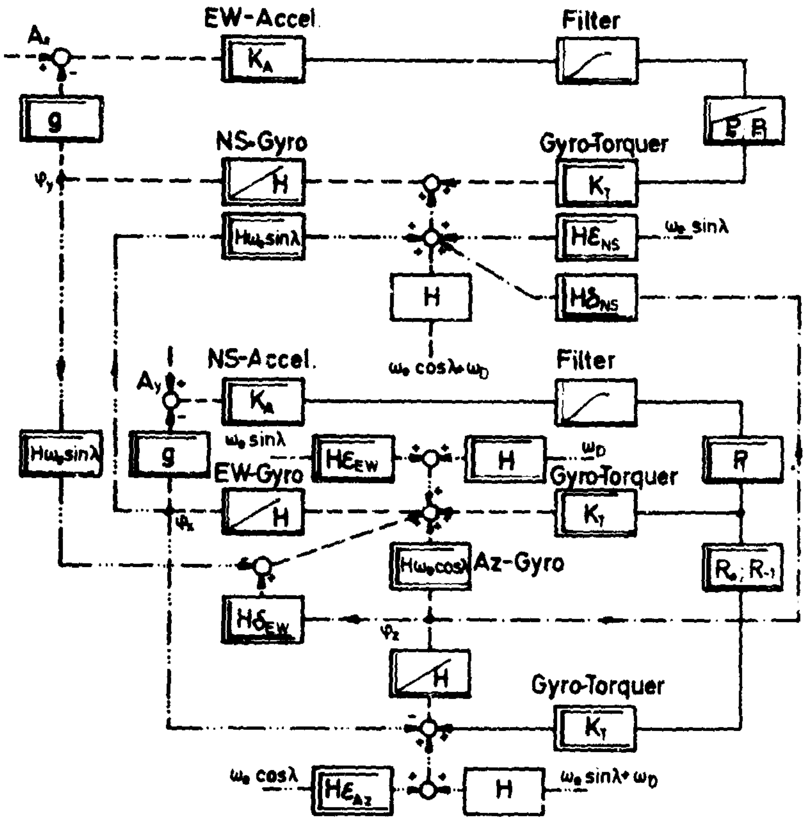


Fig.5 Block diagram of coupled alignment loops

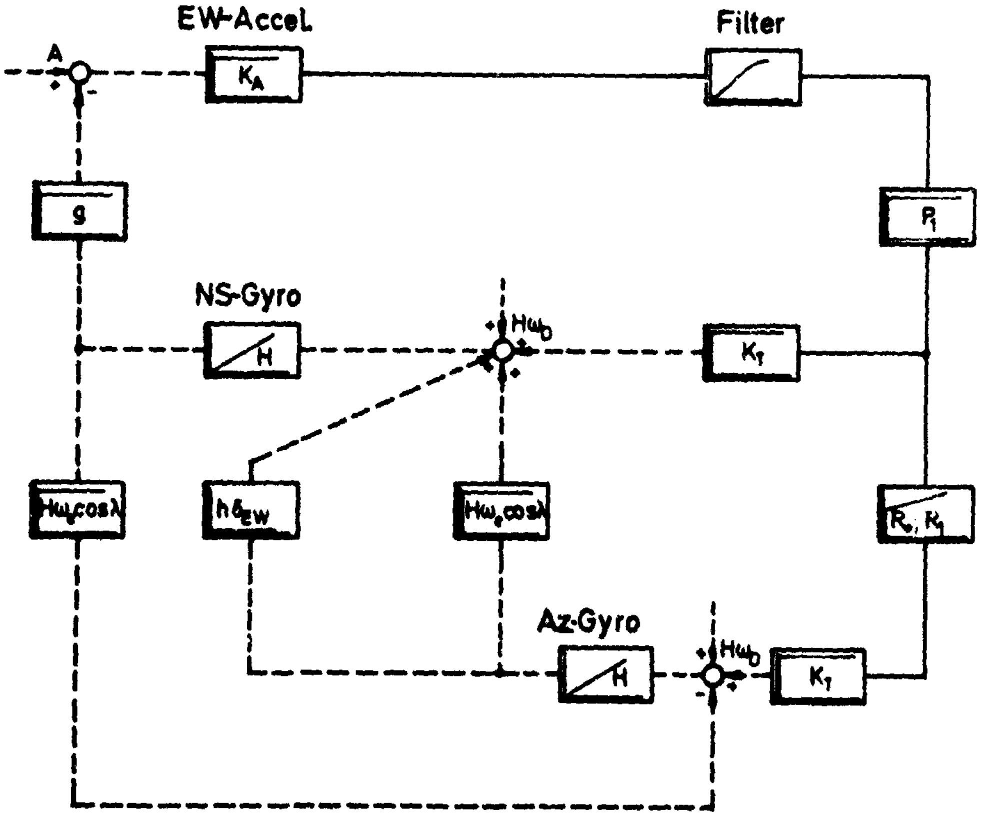


Fig.6 Effect of misalignment δ_{EW} between EW gyro and azimuth axis

PREDICTION DE L'ERREUR D'UNE CENTRALE INERTIELLE

par

Miguel da Silveira

Société d'Etudes et Conseils AERO,
3, avenue de l'Opéra, PARIS 1er, France

PREDICTION DE L'ERREUR D'UNE CENTRALE INERTIELLE

Miguel da Silveira

1. MODELES D'ERREUR

Les modèles d'erreur des centrales de navigation inertielle (CNI) interviennent aussi bien dans l'analyse des erreurs des CNI que dans les études d'optimisation de systèmes dont fait partie une CNI. Le choix du modèle est fondamental et de nombreux travaux ont été publiés où plusieurs modèles sont étudiés.

En analysant les divers modèles proposés, on s'aperçoit qu'ils font toujours intervenir des processus stochastiques *réguliers*. Les modèles basés sur des processus *singuliers* (Doob[1]) dits encore *déterministes* (Middleton[3]), ne semblent pas avoir retenu l'attention des chercheurs.

Au cours de plusieurs études effectuées depuis 1961 pour le Service Technique des Télécommunications de l'Air, et plus tard, pour la Direction des Recherches et Moyens d'Essais ainsi que pour le Service Technique Aéronautique, nous avons étudié un modèle d'erreur de CNI basé sur un processus singulier. Ce modèle a été progressivement perfectionné à mesure que des résultats expérimentaux devenaient disponibles.

2. MODELE "SINGULIER"

Rappelons que l'évolution d'un processus stochastique singulier (Doob) ou déterministe (Middleton) est défini par les valeurs d'un nombre fini de paramètres. Dans le cas des CNI, cette circonstance est à peu près vérifiée, car l'erreur de position obéit à un système d'équations différentielles; mais elle n'est pas exactement vérifiée car dans ce système, interviennent en paramètre des processus réguliers (bruits). Il se trouve que l'influence de ces bruits, au moins à court terme, est faible et on peut en profiter de la façon suivante. Si on intègre le système d'équations différentielles par la méthode de Lagrange, en le remplaçant d'abord par un système auxiliaire ne comportant pas de source de bruit, on trouve une solution générale du système auxiliaire du type

$$\vec{\zeta} = \vec{Z}(t, \vec{\sigma};\ c_1, \ldots\ldots, C_n)\ . \qquad (1)$$

où $\vec{\zeta}$ est le vecteur erreur, $\vec{\sigma}$ le vecteur position, t le temps courant et $c_1, \ldots c_n$, des constantes d'intégration. Alors la solution du système donné, comprenant les sources de bruit, est de la forme

$$\vec{\zeta} = \vec{Z}(t, \vec{\sigma}; \psi_1(t), \ldots\ldots, \psi_n(t))\ . \qquad (2)$$

où $\psi_1(t), \ldots, \psi_n(t)$ sont des fonctions convenablement déterminées, qui tiennent compte des bruits; ce seront donc des *processus*. Le fait que les bruits, en tant que processus *réguliers*, n'ont que peu d'influence, se traduit par le fait que les $\psi_i(t)$ ne s'écartent pas notablement des valeurs des constantes c_i qu'ils remplacent.

L'expression (1) de $\vec{\zeta}$ peut être considérée comme obtenue en négligeant dans (2) la variation des fonctions $\psi_i(t)$. C'est l'expression (1) qui constitue le *modèle singulier*.

La forme de $\vec{Z}$ est connue, mais les valeurs des constantes c_i sont inconnues. Par la suite, on ne s'occupera plus des fonctions ψ, c'est à dire qu'on ne s'occupera pas des bruits internes *en tant que processus*.

3. FORME EXPLICITE

En simplifiant convenablement les équations d'une CNI, on trouve l'expression approchée suivante pour le système différentiel dont $\vec{\zeta}$ est solution:

$$D^3(D^2 + [\Omega_s^2])\vec{\zeta} = [\epsilon]D^3\ddot{\vec{\theta}} \tag{3}$$

Les simplifications sont valables pour des durées et des parcours correspondant aux applications aériennes, mais pas pour des applications de grande durée. La Figure 1 indique l'erreur sur le modèle qui résulte de ces simplifications.

Dans (3), $[\Omega_s^2]$ et $[\epsilon]$ sont des (2 × 2) - matrices diagonales, la première étant scalaire. Les éléments de $[\Omega_s^2]$ ont pour valeur le carré de la fréquence angulaire Ω_s de Schüler, et ceux de $[\epsilon]$ définissent les désaccords relatifs des deux voies par rapport à Ω_s .

4. ESTIMATION DES COEFFICIENTS

La solution de (3) est constituée par deux fonctions du temps, une pour chaque voie de la centrale. Nous allons nous occuper uniquement d'une voie. La forme de la composante correspondante, ζ , est la somme de la solution générale ζ_0 de l'équation sans second membre et d'une solution particulière ζ_1 de l'équation complète, $\zeta = \zeta_0 + \zeta_1$. Soit une suite de relevés aux instants t_i : $\zeta_r(t_i)$ $(i = 0, 1, \ldots, N)$; alors ζ_0 correspond à la propagation dans le temps des erreurs existant à l'instant t_0 et ζ_1 à l'influence des accélérations du véhicule pendant l'intervalle (t_0, t) .

L'estimation de ζ_0 et ζ_1 à partir de la suite ζ_r , se fait par la méthode des moindres carrés, pour minimiser la variance d'estimation. En effet, le relevé de ζ_r est affecté de deux erreurs, celle du système de référence, par rapport auquel on évalue les erreurs de la centrale, et encore l'erreur de lecture de cette dernière.

L'expression de ζ_0 s'obtient comme une combinaison linéaire de cinq solutions linéairement indépendantes de l'équation sans second membre:

$$D^3(D^2 + \Omega_s^2)\zeta_0 = 0 . \tag{4}$$

Par exemple, on pourrait prendre comme système fondamental le suivant,

$$\cos \Omega_s t, \quad \sin \Omega_s t, \quad 1, \quad t, \quad t^2 , \tag{5}$$

mais il présente, du point de vue de l'estimation, le défaut d'être oblique. Il est préférable de prendre cinq fonctions orthogonales, que nous prenons aussi normées, soit $\xi_1(t), \xi_2(t), \ldots, \xi_5(t)$.

Donc
$$< \xi_h . \xi_k > = \delta_{hk} . \tag{6}$$

Le produit scalaire est défini, à un facteur constant près, par la matrice inverse de la matrice de covariances des erreurs qui affectent ζ_r . (Ref.2). En admettant que l'échantillonnage de ζ_r est suffisamment lent pour que les erreurs deviennent décorrélées, et que leur variance est constante, on peut poser

$$< \xi_h . \xi_k > = \sum_{i=0}^{N} \xi_h(t_i)\, \xi_k(t_i) . \tag{7}$$

En partant du système fondamental (5) et en lui appliquant la méthode d'orthonormalisation de Schmidt, avec le produit scalaire (7), on trouve une base orthonormée $\{\xi\}$.

Les avantages statistiques de l'emploi des bases orthonormées sont bien connus : non corrélation des valeurs estimées, facilité de l'analyse de variance, remplacement de la distribution de Fisher-Snedcor par celle de Student, qui est robuste, etc. Du point de vue numérique, on évite les inversions de matrices, et un même système de coefficients permet de trouver tous les ajustements partiels[2].

Nous avons encore profité de la souplesse qui résulte des bases orthonormales pour compléter $\{\xi\}$ par un terme d'alarme ξ_6 . Ce terme est choisi de façon à ne pas être solution du système (4). Si l'analyse de variance conduit à lui attribuer un coefficient significatif, cela indique que l'intervalle (t_0, t_N) est trop long.

Les coefficients c_h des ξ_h s'obtiennent à partir de la composante ζ_{0r} de ζ (qui correspond à ζ_0 dans ζ) par les formules classiques

$$\hat{c}_h = \langle \xi_h, \zeta_{0r} \rangle , \tag{8}$$

leurs variances étant égales entre elles et égales à celle des erreurs de mesure.

L'estimation des coefficients de ζ_0 présuppose que l'on a pu préalablement décomposer ζ_r en ζ_{0r} et ζ_{1r} . Pour ce faire, nous employons la méthode des "fonctions modulantes"[3]. En effet, de l'équation qui définit ζ , on peut déduire ϵ , en appliquant aux deux membres une fonctionnelle linéaire continue convenable, que nous représentons au moyen de la formule de Riesz. On obtient ainsi l'expression suivante pour l'estimateur de ϵ :

$$\hat{\epsilon} = \frac{\langle D^3 (D^2 + \Omega_s^2)\alpha, \zeta_r \rangle}{\langle D^5 \alpha, \vartheta \rangle} , \tag{9}$$

où α est une fonction qui satisfait certaines conditions aux limites. C'est α qui constitue la "fonction modulante"[3,4].

On choisit α par la condition de minimiser la variance de $\hat{\epsilon}$, ce qui conduit à l'équation différentielle du 10ème ordre sur α ,

$$D^6 (D^2 + \Omega_s^2)^2 \alpha = D^5 \vartheta , \tag{10}$$

avec les conditions aux limites suivantes:

$$\left.\begin{array}{l} \alpha(t_0) = \alpha'(t_0) = \ldots\ldots = \alpha^{(4)}(t_0) = 0 \\ \alpha(t_N) = \alpha'(t_N) = \ldots\ldots = \alpha^{(4)}(t_N) = 0 \end{array}\right\} . \tag{11}$$

Connaissant l'estimation $\hat{\epsilon}$ du déréglage, on obtient l'estimation $\hat{\zeta}_{1r}$ de ζ_{1r} en intégrant l'équation avec second membre qui définit ζ . En introduisant les conditions aux limites convenables, on trouve

$$\hat{\zeta}_{1r} = \frac{\hat{\epsilon}}{\Omega_s} \Big[H(t-t_0) \sin \Omega_s (t-t_0) \Big] * \Big[H(t-t_0) H(t_N - t) \vartheta(t) \Big] \tag{12}$$

où H désigne l'échelon unité et $*$ la convolution. Dans ces conditions, l'estimation de ζ_{0r} s'obtient par soustraction:

$$\hat{\zeta}_{0r}(t) = \zeta_r(t) - \hat{\zeta}_{1r}(t) . \tag{13}$$

C'est cette fonction $\hat{\zeta}_{0r}$ qui permet de calculer les coefficients c_h par (8).

On obtient finalement la représentation suivante de l'erreur ζ :

$$\hat{\zeta}(t) = \sum_{h=1}^{5} \hat{c}_h \tilde{\xi}_h(t) + \hat{\zeta}_{1r}(t) . \tag{14}$$

On démontre que la variance d'estimation de $\hat{\zeta}(t)$ est donnée par

$$\frac{\sigma^2_{\hat{\zeta}(t)}}{\sigma^2_r} = \sum_{h=1}^{5} \tilde{\xi}_h^2 + \frac{\langle D^3(D^2+\Omega_a^2)\alpha, D^3(D^2+\Omega_a^2)\alpha\rangle}{\langle D^5\alpha,\sigma\rangle^2}\,\dot{\zeta}^2_{1r}(t) , \tag{15}$$

σ^2_r étant la variance des erreurs de mesure.

5. RESULTATS

Les idées générales qui viennent d'être rapidement exposées ont été dégagées peu à peu au fur et à mesure que les données expérimentales devenaient disponibles en France. C'est grâce aux remarquables campagnes d'essais entreprises par le Service Technique Aéronautique que nous avons pu éprouver la validité des conceptions théoriques.

5.1

Ainsi, un essai préliminaire effectué en 1964 a permis une première vérification d'un modèle singulier ne comprenant pas la partie ζ_1 due aux accélérations du véhicule. La Figure 2 montre le résultat obtenu. L'ajustement s'étendait sur une durée de 2000 s, l'échantillonnage se faisant à la cadence d'un point toutes les 100 s. Les points employés pour l'estimation des coefficients portent les numéros 0 à 20, et les points suivants correspondent à l'extrapolation. Pour chaque point à partir du 21ème, la croix indique la valeur mesurée et les cercles les valeurs extrapolées. Le chiffre indiqué au-dessus donne le rang du dernier terme du développement employé. L'analyse de variance indique que les termes de rang 1, 2, 3 et 4 sont significatifs mais celui du rang 5 et le terme d'alarme (range 6) ne le sont pas. On voit que l'extrapolation est nettement améliorée quand on ajoute le terme de rang 4. Le terme de rang 5 (non significatif) n'ajoute rien. Mais le fait remarquable, c'est que si on introduit le terme de rang 6, *l'extrapolation devient aberrante*. Ce phénomène est dû à ce que la fonction $\tilde{\xi}_6$ *n'est pas* solution de l'Equation (4).

Cet exemple est intéressant, car la fonction $\tilde{\xi}_6$ a été obtenue en ajoutant à la suite (5) un terme cubique t^3. Or, on peut considérer que les termes 1, t, t^2 et t^3 constituent le début d'un développement de Taylor qui permettrait de serrer la réalité de plus près à mesure que des termes sont rajoutés. Ce serait le cas si l'erreur ζ correspondait à un modèle *régulier*, car il ne saurait pas être défini par une équation différentielle à coefficients certains. Or, on voit que l'augmentation du degré du développement n'apporte aucune amélioration, et que si on le pousse au degré 3 ($\tilde{\xi}_6$), il devient aberrant. Donc, le terme en t^2 (apporté par $\tilde{\xi}_5$) n'est pas incorrect, mais celui en t^3 est complètement faux. Cette circonstance a montré que l'évolution de l'erreur ζ obéit bien au modèle singulier proposé.

5.2

Depuis, d'autres essais ont été réalisés, qui ont montré le besoin de tenir compte de l'accélération du véhicule. C'est ce qui a conduit à introduire le terme ζ_1. Ce modèle plus complet n'a pas encore été essayé*. Les essais réalisés montrent que lorsque le mouvement est à peu près rectiligne et uniforme, l'extrapolation est excellente. Mais lorsque des accélérations sont présentes, l'extrapolation devient immédiatement fausse.

---

* Des essais postérieurs au Symposium AGARD ont été effectués. Nous donnons en additif les résultats déjà obtenus avec le modèle (14) complet.

Les courbes de la Figure 3 montrent des exemples d'ajustement sur des segments où le mouvement du véhicule était à peu près uniforme, suivis de virages. On voit que l'extrapolation est correcte jusqu'au moment des virages, mais qu'après l'erreur devient considérable.

Les courbes de la Figure 4 montrent ce qui arrive quand l'intervalle d'ajustement contient des accélérations. L'estimation conclut à l'existence de termes en t^2 (coefficient de ξ_5 significatif), alors que ce terme est pratiquement absent. Ceci est dû aux accélérations qui, à travers le terme ξ_1 dont il n'a pas été tenu compte, déforment la courbe $\xi(t)$.

On voit ainsi qu'il est nécessaire de tenir compte de l'influence de l'accélération.

Les résultats expérimentaux analysés jusqu'à présent, bien que ne faisant pas usage du modèle complet (14)*, ont montré que la plus grande partie des erreurs provient du désaccord de la centrale et non de l'éloignement initial. C'est en analysant ces résultats que nous avons été conduits à compléter ξ_c par le terme ξ_1 .

6. ADDITIF

Des essais récents ont été effectués après l'introduction du terme ξ_1 de déréglage de période. Un exemple caractéristique est fourni à la Figure 5 : on peut y constater une variation rapide de la pente des erreurs à l'instant d'un virage brutal effectué par l'avion; un phénomène du genre ne peut évidemment pas être décrit par les termes du modèle ξ_c , et il correspond à un déréglage de période de 1,2%. Les valeurs prises par ξ_1 après ce virage atteignent 8 km, et montrent l'importance de ce terme.

L'ensemble des résultats obtenus avec le modèle (14) complet a toutefois mis en évidence la présence d'un résidu d'erreur fonction de l'accélération du véhicule, ce qui conduit à envisager l'influence de termes tels que les balourds des gyroscopes de verticale et la dérive du gyroscope d'azimut. Ceci confirme l'opinion exprimée par le Dr Frey que le modèle (14) était incomplet et qu'il faudrait tenir compte des balourds.

REFERENCES

1. Doob, J.L.	*Stochastic Processes* John Wiley, New York, 1953.
2. Kendall, M.G. Stuart	*The Advance Theory of Statistics* Vol.2, Charles Griffith, 1961.
3. Middleton, D.	*Statistical Communication Theory* McGraw-Hill, New York, 1960.
4. Loeb, J. Cahen, G.	*Extraction à Partir des Enregistrements de Mesures des Paramètres Dynamiques d'un Système* Automatisme, Vol.8, No.12, Décembre 1963.
5. Loeb, J. Cahen, G.	*More About Process Identification* Transactions, Institute of Electrical and Electronic Engineers, AC-10, No.8, July 1965.

---

* Voir note page 114

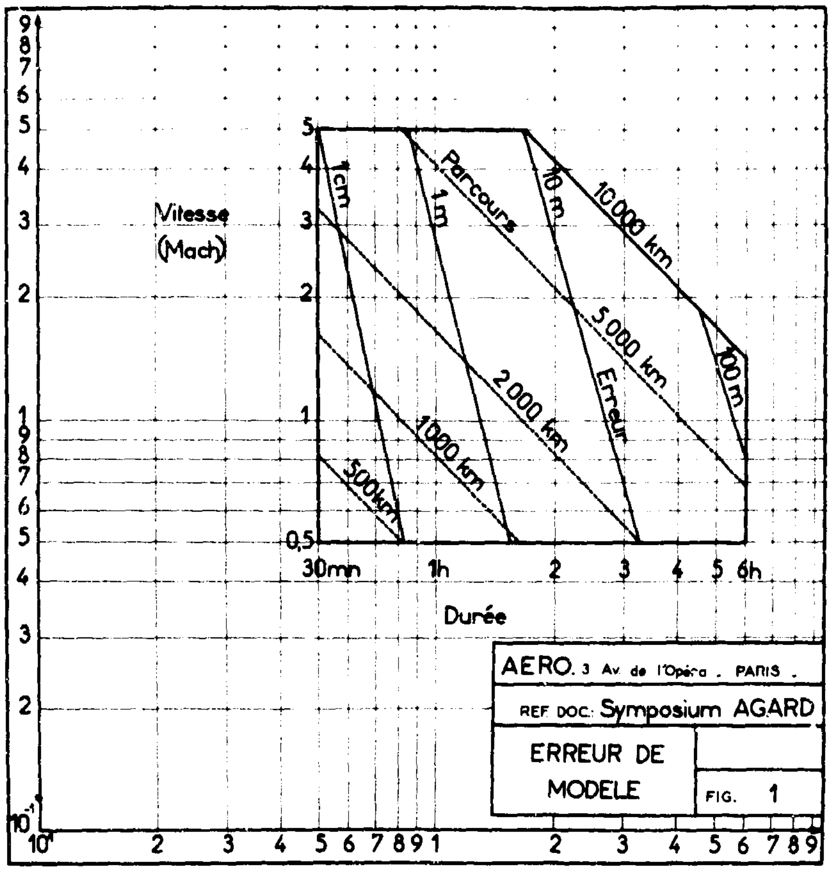


Fig. 1 Erreur de modèle

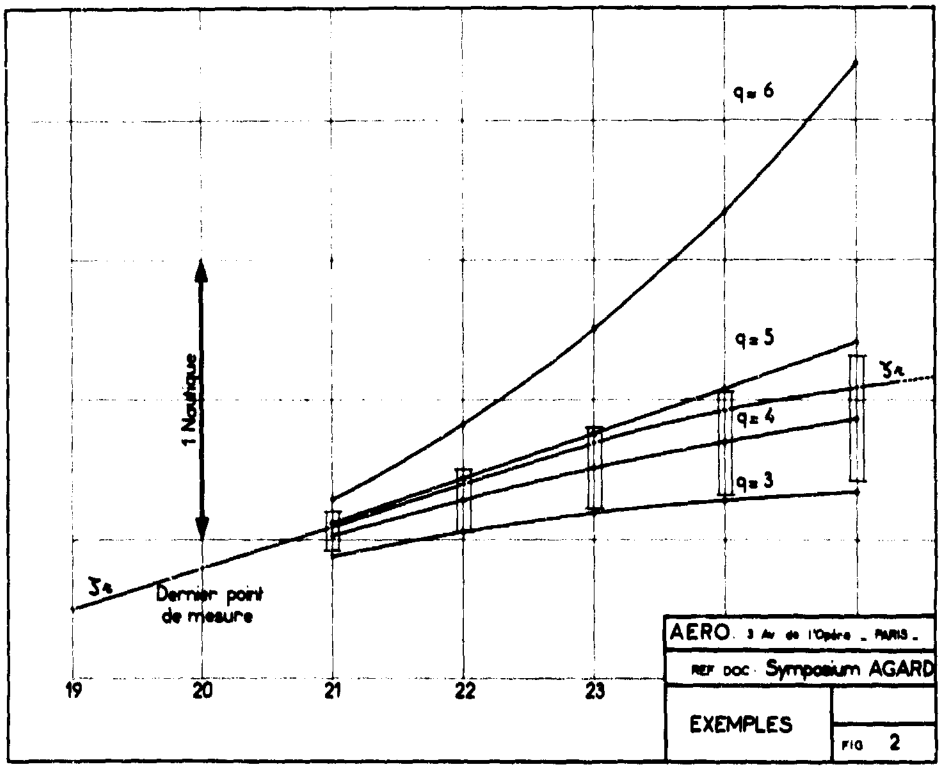


Fig. 2 Exemples

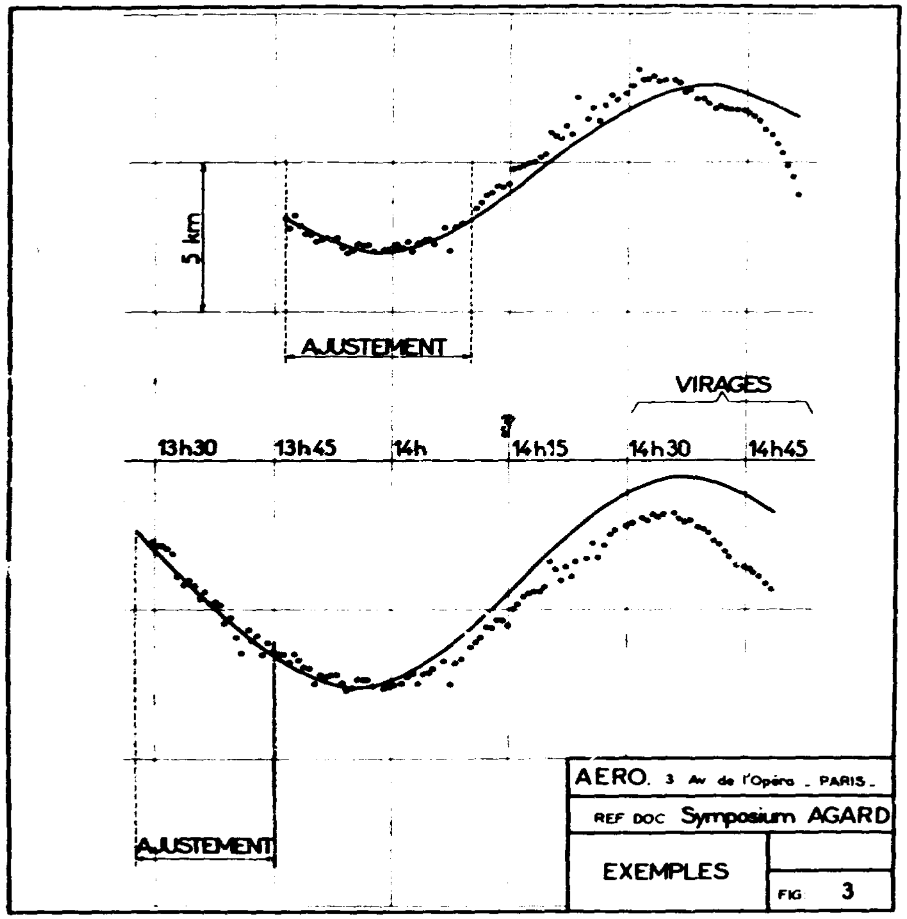


Fig.3 Exemples

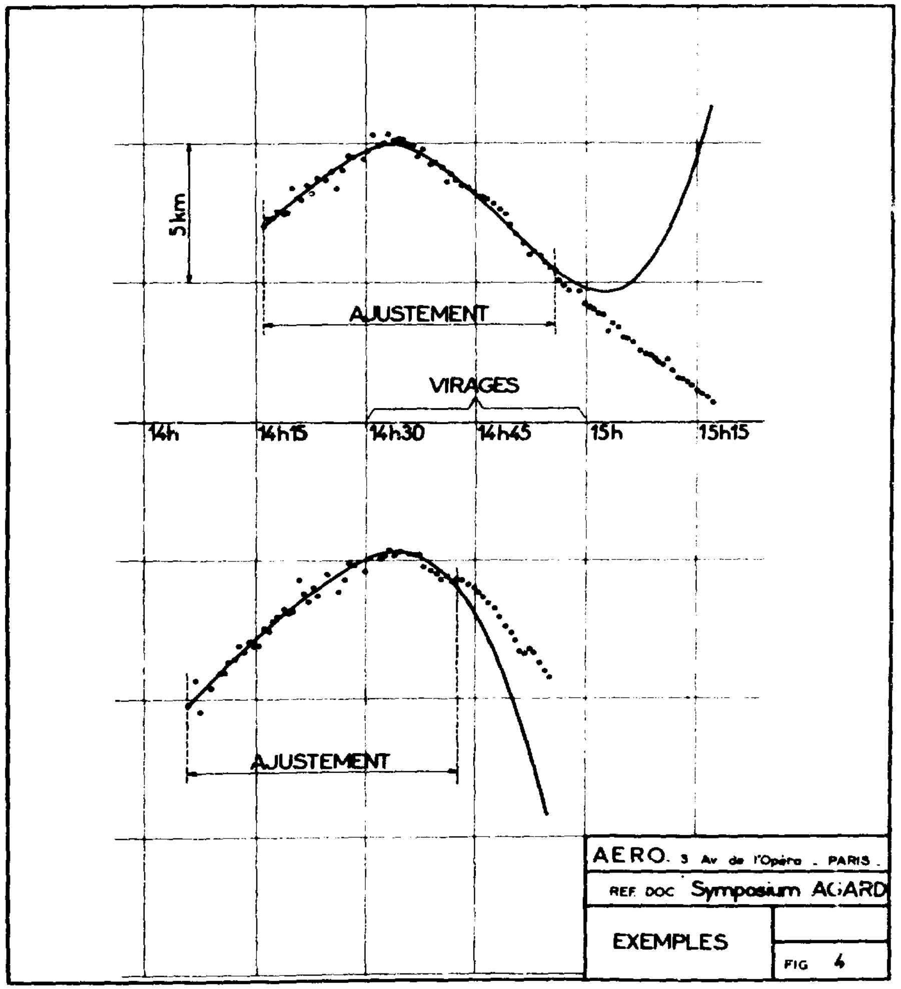


Fig.4 Exemples

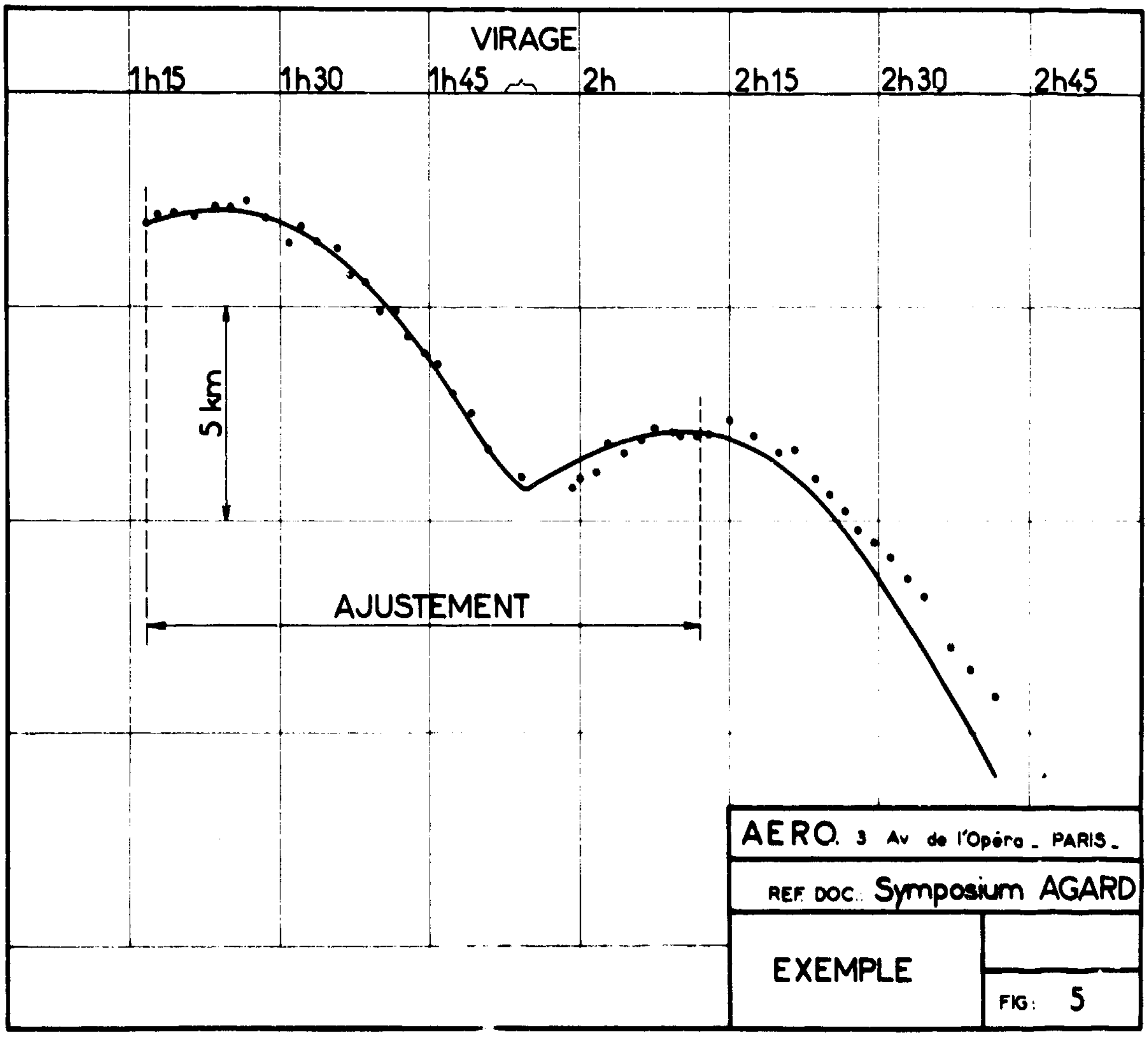


Fig.5 Fxemple

FACTORS INFLUENCING THE CHOICE OF COMPUTER FOR AN INERTIAL NAVIGATION SYSTEM

by

P.Wilson

Elliott Brothers (London) Limited,
Airport Works, Rochester, Kent, England

SUMMARY

The paper deals exclusively with digital computing techniques. The dominance of digital computing techniques in the inertial field is now almost complete and a resumé is made of the reasons leading to this situation.

However, there is a very wide range of digital computing techniques available and for the purpose of discussion these are divided into three broad classifications:

(i) General purpose computers
(ii) Digital differential analysers
(iii) Special purpose computers

Each of the classifications is taken in turn and the merits and disadvantages are discussed. The technical merits of the computing system itself, however, are not the only consideration, and many other influences may affect the final choice.

Finally, a survey is made of some typical systems and the probable reasons leading to their particular configuration. An attempt is made to forecast the future trend in the light of present day inertial system requirements.

FACTORS INFLUENCING THE CHOICE OF COMPUTER FOR AN INERTIAL NAVIGATION SYSTEM

P.Wilson

1. INTRODUCTION

The dominance of digital computing techniques in the inertial field is now virtually complete and this paper deals exclusively with the various digital techniques available.

The overwhelming advantage of the digital computer is, of course, the achievement of consistently high computing accuracy over extended periods of time without the inherent errors which necessarily occur in analogue computing such as variations in scale factor, calibration drifts and similar instrumental errors. It is significant that, whereas the first inertial navigation systems used in missile applications having relatively short flight times could satisfactorily rely upon analogue computing techniques, as these systems have become more widely applied to cruise vehicles, digital computing has steadily replaced the analogue approach. Indeed it is probable that without the inherent and constant computing accuracy available with digital techniques, it would not have been possible to mechanise inertial navigators for modern day long-range aircraft without excessive penalties in weight and cost.

From a practical point of view, in the wide-scale application of inertial systems, the relative freedom from maintenance problems and the improved reliability given by the digital approach is a factor of great importance.

Of nearly equal importance is the ability to use the digital computer to mechanise complex mathematical relationships. The most obvious example is the use of the digital computer to provide a trans-polar flight capability by one of several different means, one of the most widely used being the wander azimuth mechanisation. Another example is the compound navigation system in which the DR position derived by the inertial systems is continuously compared with, and sometimes corrected by, the position outputs of other navigational aids.

In considering the various digital techniques available it is convenient to divide them into three broad classifications:

(i) General purpose computers.
(ii) Digital differential analysers.
(iii) Special purpose computers.

2. DIGITAL COMPUTING TECHNIQUES

2.1 The General Purpose Computer

In this context "general" implies a generalised method of handling the IN calculations, rather than the use of a computer suitable for instant application to other problems.

This type of computer makes use of a central processing section capable of arithmetic operations and certain decision functions, and under the control of a stored instruction sequence. A simplified block diagram is shown in Figure 1. In this type of computer, calculations are performed on whole numbers, usually represented in binary notation, and calculations are performed sequentially according to the instruction sequence. Where calculations are performed on basically analogue or incremental data and smoothly varying outputs are required, the input data must first be calculated to whole number form, and the calculations must be repeated at a sufficiently frequent interval to give the output resolution required.

There are three main variables in the design of a general purpose computer which are relevant to its choice for an IN system. Firstly, the number of digits in the computer word has a direct effect on the calculation accuracy. For most IN applications a wordlength of 18 binary digits is adequate but various expedients can be used to extend the accuracy of calculation, thus allowing a shorter basic wordlength to be used. Secondly, the capacity of the computer memory governs the length of the programme (instruction sequence) which can be stored and the quantity of present data available to the programme. Additional capacity can usually be provided quite cheaply and may tempt the designer to extend the use of the computer to functions other than navigation; however, other considerations, such as integrity, usually rule against this. A memory capacity of 2000 to 3000 instructions is usually sufficient for a typical navigation problem. Thirdly, the speed of the computer is related to the repetition frequency of the instruction sequence and to the number of instructions which may be obeyed in that sequence in a given time. Unless other functions are to be performed by the computer, the speed requirements for inertial navigation are usually quite low and seldom present a major problem.

From the inertial system designer's viewpoint one of the main advantages of the GP approach is that he will probably have available to him an existing design of computer with a known performance and reliability record and will thus be able to avoid the heavy financial burden associated with the design and development of a computer special to his requirements. This advantage may, however, be offset by the fact that as a rule most GP computers have a store capacity of 8000 words or more, which is significantly in excess of the 2000 to 3000 words required for the basic navigation functions. There is naturally a higher manufacturing cost associated with this size of computer and it is therefore a matter for careful judgement as to which is the most economical solution overall. In recent years the tendency has been for the requirements of the IN system to increase (storage of waypoints, computation of great circle course and distance to go, steering outputs to autopilot) which favours the GP machine. Indeed, this ability to cope with changes and additions to the original requirement specification is perhaps the second most important advantage of this class of computer.

A disadvantage often encountered is the sometimes extensive conversion circuits necessary to convert output signals to the required electrical form (usually analogue). This problem can often be avoided if the computer is tailor-made for the application, as is the case with the special purpose type of computer. It is expected that, with the wider application of digital techniques to other avionic equipment, this particular problem will disappear.

2.2 The Digital Differential Analyser

Unlike the general purpose computer, the digital differential analyser grows with the size of the problem. In operation, its major difference is that it computes incremental changes to quantities rather than changes to the whole values of the quantities themselves. Its use, therefore, is limited almost exclusively to sums on continuous variables and, because of the method of mechanisation, to time-dependent variables.

The basic computing process of a DDA is one of counting, or of incremental integration, as shown in the block diagram (Fig.2). In mathematical terms, a problem is broken down into equations of the form $dz = ydx$, where x, y, and z are all time-dependent variables, and incremental quantities are represented by pulse trains synchronised with the operating cycle of the computer. Addition and subtraction of pulse trains are possible;

multiplication and division can be performed by mathematical manipulation. Some simple decision functions are usually available, but they are much more clumsy in operation than the equivalent functions in a general purpose computer.

There are two main types of DDA, the parallel and the serial. The parallel DDA has separate hardware for each incremental integration performed, and is therefore very nearly the digital equivalent of the conventional analogue computer. Its logic is fixed by means of physical interconnection of integrators and it is rather rigid in its operation. The serial DDA is more closely related to the general purpose machine. The parts of the serial DDA concerned with arithmetic and control are time-shared in a sequential series of operations on stored data. In one cycle, one complete iteration would be performed. The sequence of instructions is usually stored along with the data, and in some machines the interconnection of the integrators may be stored likewise. This makes its operation more flexible than that of the parallel DDA, but nevertheless it is more clumsy than the general purpose computer and is strictly limited in the type of problem it can solve.

The word length and speed of operation of a DDA are related to the accuracy and smoothness of output in much the same way as the general purpose computer. However, besides smoothness, one must consider the rate of change of variables and ensure that the maximum rate of change expected is within the speed capability of the DDA. This is because transitions from one value to another must be based on incremental changes of limited step size and frequency, as compared with the GP computer which recalculates on the basis of the whole number value of new data each cycle.

The main value of a DDA is its ability to tackle small problems efficiently and its convenience when used as an extrapolation device where simplified functions within the capability of the DDA can be used over very limited periods. Its disadvantages are the difficulty in programming large problems except on a very sophisticated machine, its inflexibility, and drift problems arising from the impossibility of representing exactly many commonly used functions by incremental equations.

In summary, therefore, the DDA type of computer is suitable only for those applications where the system requirements in terms of facilities and outputs are small in number and not likely to change. In view of earlier comments on the tendency for requirements to increase and the desire of the system designer to retain flexibility to cope with additions to the initial requirements, the DDA approach is losing most of its former popularity.

2.3 The Special Purpose Computer

Any computer designed to perform a particular function can be considered to be a special purpose computer. However, here the term is used to describe computers that are neither purely general purpose nor purely digital differential analysers, but may include some of these techniques and others, including counting and pulse modulation.

It is not proposed to describe any particular special purpose computer, since each design is likely to employ different principles.

The advantages of the special purpose computer are the fact that it can be optimised both overall and in local areas of calculation to employ the most suitable computing techniques, accuracy and speed for that particular part of the problem. This can take into account the form of input data and types of output concerned and will minimise specialised interface circuitry required.

However, the adoption of a special purpose solution results in a non-standard device, both at unit and at computing element level. This brings with it economic disadvantages compared with an equivalent size standard machine and inherent difficulties in maintenance and testing.

The special purpose computer therefore presents a good solution to the problem from a cost-effective viewpoint in cases where the computational problem is not too complex. A good example of this is a system which requires only outputs of latitude and longitude. As the requirement for additional outputs and facilities is increased, this advantage rapidly diminishes. The summary applied to the DDA class of computer is therefore appropriate here also, and the lack of flexibility of the special purpose machine is probably more restrictive than that of the DDA.

3. FUTURE TRENDS

It already seems clear that the size of the IN computing problem (ARINC 561 being a good example) has reached the point where the DDA and special purpose computer approach can show little or no cost advantage over the GP machine (see Figure 3). From its very nature the general purpose computer can be expected to find wider application in avionics as well as in other fields and this will do much to reduce both manufacturing costs and amortised development charges. With the growing use of digital techniques within other avionic systems, the general purpose computer offers an increasingly useful method of distributing, combining, and comparing information between different parts of the aircraft in a common signal format. The advantages of this rationalisation of signal transmission have already been recognised and valuable work is going on within ARINC to attempt to establish a common digital signal transmission format for an increasing number of types of information within civil aircraft. Application of this technique will permit very significant simplifications and weight reductions in aircraft cabling.

In addition, the greatly superior flexibility provided by the general purpose computer is of inestimable value. It provides the ability to accommodate changes to the requirement specification, it enables the designer to supply basically the same hardware to applications having different requirements and, from the overall systems aspect, it makes possible the engineering of redundancy modes in a very practical way.

For the reasons already outlined, it seems likely that future inertial navigation systems will concentrate on general purpose computers. For reasons of integrity, it does not seem likely that these computers will be large single units, with an extensive memory capable of carrying out many other functions additional to the navigation system. Such a system concept poses many organisational problems, even if the safety requirement is met by a duplicate computer installation. Furthermore, with large computers of 16,000 words or more capacity, the system capability is no longer limited by the storage capacity of the computer but tends to be restricted by the time taken up in the computing cycle. Programming techniques have been developed to alleviate this problem, but they involve a more complicated programme which is difficult to change in the light of operational experience.

In the past, the high cost of general purpose computers has necessitated their use for a number of different tasks in order to "justify" their cost. However, the rapid reduction in computer costs, arising from their wider application and from the introduction of microelectronics now makes it possible to manufacture "single task" GP computers at costs competitive with the "large" computer system.

Such single task "minimal" computers will have a relatively small storage capacity of around 4000 words, and will have a short word length around 12 bits, employing double length working where higher accuracies are required. Because the computer will be restricted to a single task, the speed limitations of the larger computer should not intrude and a $2\mu s$ parallel store will probably be quite adequate.

The future envisaged is that a standard general purpose computer in this "minimal" form will be the basis of most inertial navigation systems. It is expected that the same machine will then find wider application to other single task functions such as air data systems, autopilots, engine control and general flight management problems.

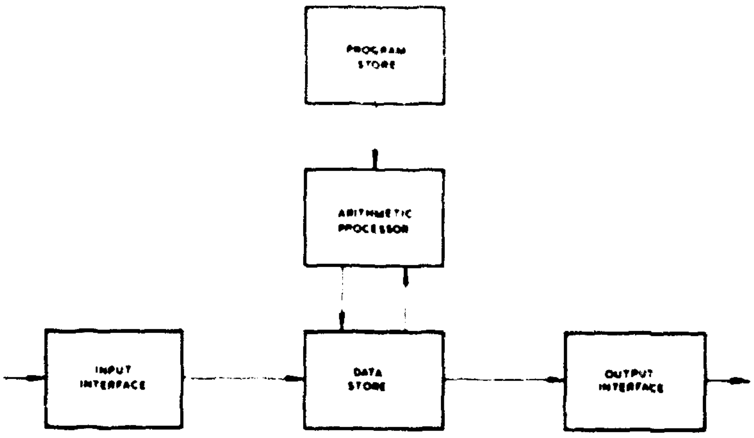


Fig. 1 Block diagram of a simple general purpose machine

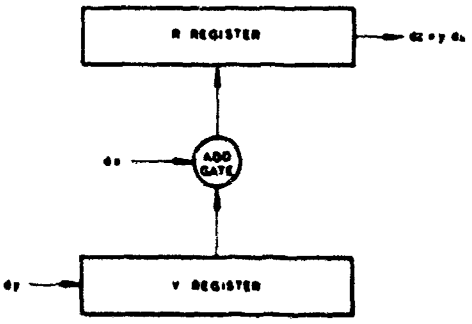


Fig. 2 Basic DDA integrator

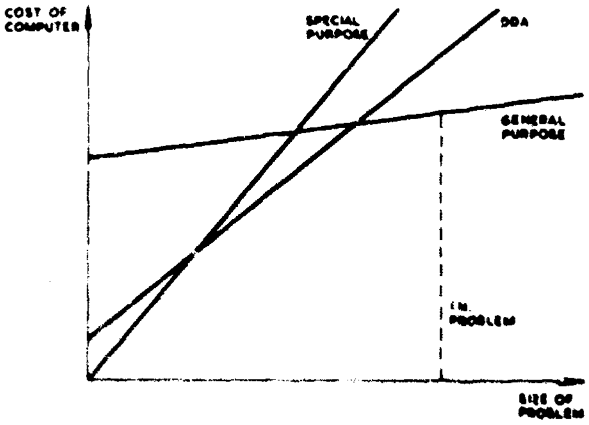


Fig. 3 Single computer economics

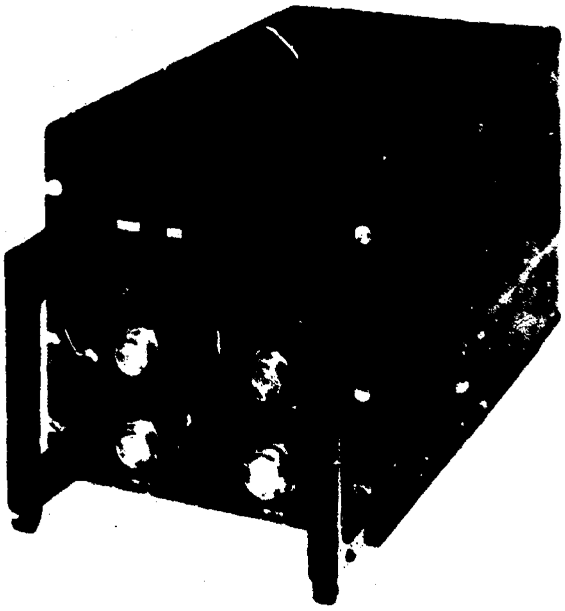


Fig. 4 The MCS920M digital computer for airborne control

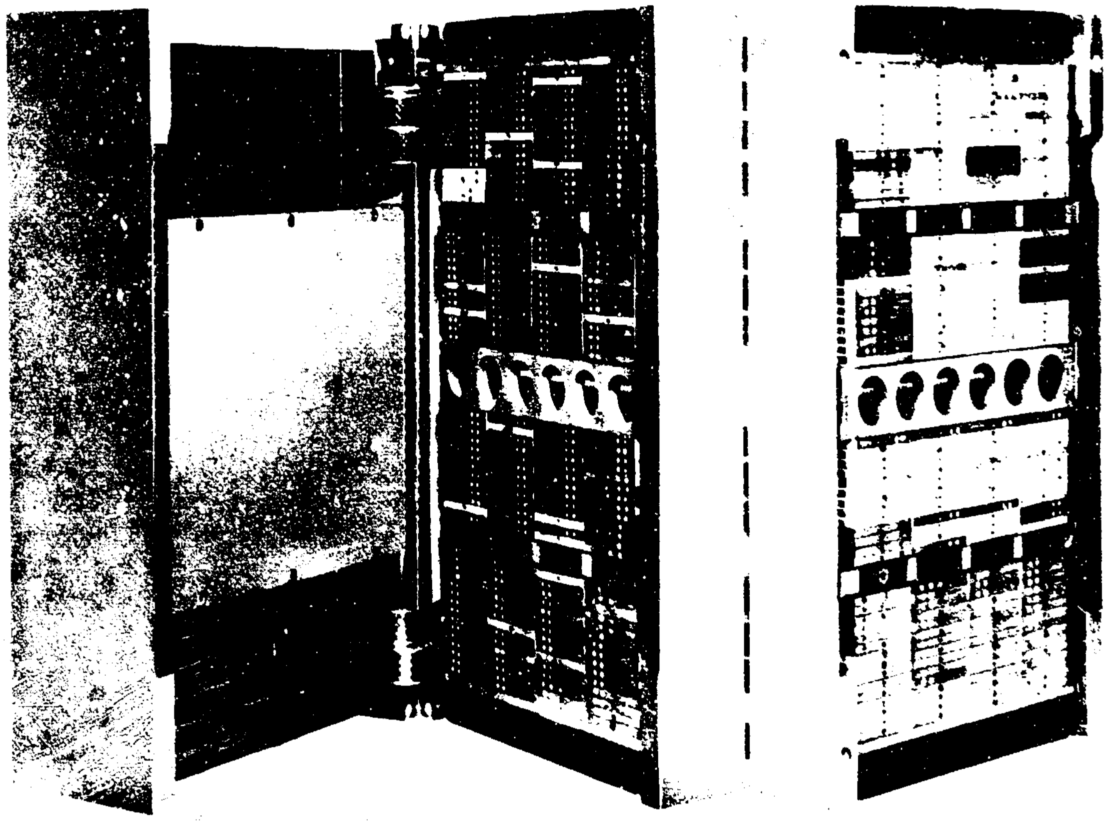


Fig. 5 The MCS920 computer, hinged open

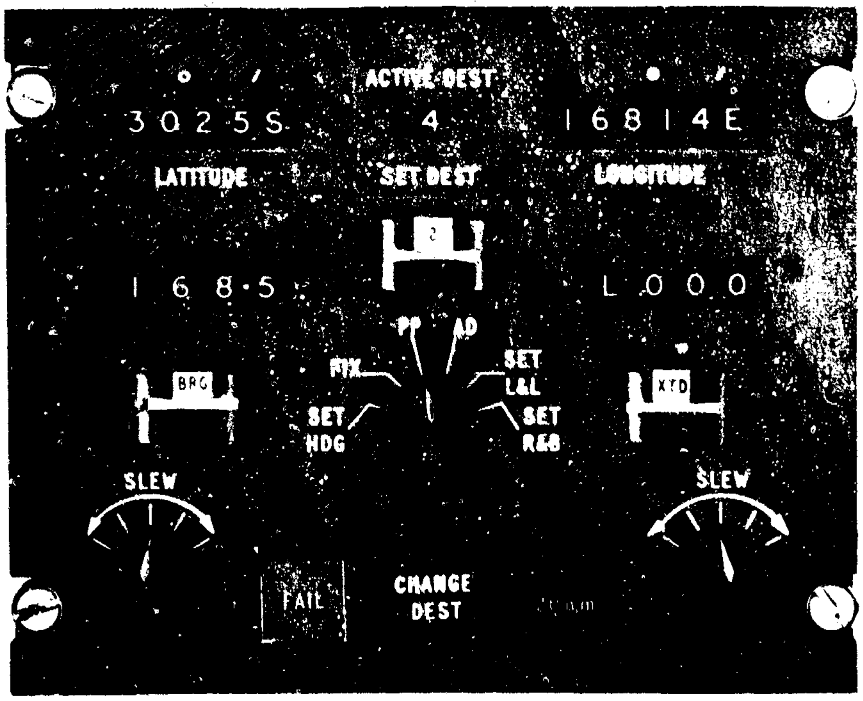


Fig. 6 Airborne display unit

OPTIMAL USE OF REDUNDANT DATA IN INERTIAL NAVIGATION SYSTEMS

by

F.G.Unger

Teldix Luftfahrt-Ausrüstungs GmbH,
Heidelberg, Germany

SUMMARY

The performance of inertial navigation systems can be improved by using additional or redundant data from various sensors. Several methods are available to update the inertial navigator, reaching from pure resetting to optimal filtering. These methods, their possible implementations, their different influences on the behaviour of the system, and the resulting errors are described.

OPTIMAL USE OF REDUNDANT DATA IN INERTIAL NAVIGATION SYSTEMS

F.G.Unger

1. INTRODUCTION

Accurate and reliable navigation is a very stringent requirement for military aircraft. The required outputs from a navigation system are position, velocity and orientation of the sensors and the vehicle with respect to a navigational coordinate system. A pure inertial navigation system can provide all this required information, but the errors of an inertial system are not bounded, i.e. they increase with time. Therefore, with large mission times, the errors can become intolerably large. Other methods of navigation, like Doppler dead reckoning, radio, stellar etc. can provide only part of the information required, but with bounded errors in velocity or position. A navigation system that fulfils a wide variety of requirements for relatively long mission times will, therefore, be a combination of an inertial system with other navigational aids.

This paper presents the different methods of aiding a pure inertial system with additional or redundant navigation information. The available sensor outputs are categorised and a general block diagram for navigation systems is derived. The pure inertial system in a Schüler-tuned mode is included in this block diagram as a special case. The use of redundant information in a conventional way, by proportional or integral feedback, is then discussed. The last section describes the use of a Kalman or minimum variance filter to update the navigation loops.

2. GENERALISED BLOCK DIAGRAM OF NAVIGATION SYSTEMS

The central element of a navigation system is the navigation computer. Its inputs are data from sensors, such as inertial acceleration, ground speed, air speed, heading, position information, stellar angles, etc. These data may be continuous, in analogue or digital form, or intermittent, like ground fix points. The computer has to calculate the required navigation outputs: position, velocity and orientation of the sensors and the vehicle with respect to a navigational coordinate system. In addition the computer must also compensate for sensor biases and calculate torquing commands for platform gyros or positional servos for stellar or radar equipment.

Figure 1 shows the possible inputs and required outputs of a navigation system.

The sensor information available, as well as the computed output data, can be classified into three levels. They are:

Level 1: Acceleration
Level 2: Velocity
Level 3: Position and Orientation.

Orientation belongs to the same level as position, because angular measurements lead to position information and both data complement each other. Furthermore, in a rotating coordinate frame, orientation angles can be derived by integration of the angular rate vector $\vec{\omega}$, just as position is derived by integration of the velocity vector $\vec{v}$.

Figure 2 shows the three levels and the equivalent inputs.

Information for a higher level can be derived from a lower level by integration; but it is generally not possible, to derive lower level information from higher levels by differentiation.

The information in each level is a vector in the navigation coordinate frame except for orientation, where the information is in the form of three angles or the direction cosine matrix (M) for possible coordinate transformations.

If the navigation frame is rotating with respect to inertial space, corrections for acceleration must be included. This can be done by writing the basic equations as

$$\left(\frac{d^2\vec{R}}{dt^2}\right)_I = \vec{a} + \vec{g}_m \tag{1}$$

$$\left(\frac{d\vec{R}}{dt}\right)_I = \vec{v} + \vec{\omega} \times \vec{R} \tag{2}$$

$$\left(\frac{d^2\vec{R}}{dt^2}\right)_I = \left(\frac{d\vec{v}}{dt}\right)_N + \dot{\vec{\omega}} \times \vec{R} + 2\vec{\omega} \times \vec{v} + \vec{\omega} \times (\vec{\omega} \times \vec{R}) \ , \tag{3}$$

where $\vec{R}$ = position vector

$\left(\frac{d\vec{R}}{dt}\right)_I$ = time derivative of $\vec{R}$ with respect to inertial space

$\left(\frac{d\vec{R}}{dt}\right)_N$ = $\vec{v}$ = time derivative of $\vec{R}$ with respect to the navigational coordinate frame

$\vec{\omega}$ = rate of change of navigational coordinate system with respect to inertial space

$\vec{a}$ = thrust acceleration vector

$\vec{g}_m$ = vector of mass attraction

$\vec{b}$ = $\left(\frac{d\vec{v}}{dt}\right)_N$ = total acceleration in navigation coordinates.

Substitution of Equation (3) into Equation (1) yields

$$\vec{b} = \vec{a} + \vec{g}_m - \dot{\vec{\omega}} \times \vec{R} - 2\vec{\omega} \times \vec{v} - \vec{\omega} \times (\vec{\omega} \times \vec{R}) \ . \tag{4}$$

Equation (4) shows that the following corrections are required on the acceleration level in addition to the thrust acceleration:

1. Acceleration due to mass attraction: $\vec{g}_m$
2. Coriolis acceleration : $2\vec{\omega} \times \vec{v}$
3. Tangential acceleration : $\dot{\vec{\omega}} \times \vec{R}$
4. Centripetal acceleration : $\vec{\omega} \times (\vec{\omega} \times \vec{R})$.

If the sensitive axes of the sensors do not coincide with the axes of the navigational coordinate system, the sensor outputs must be transformed, using the orientation matrix (M).

Figure 3 shows a general block diagram of navigation systems including the corrections for a rotating frame. Redundant data may enter the system at levels 2, 3a and 3b. The problem of how the redundant data are used within the level blocks is discussed in Section 4 and 5.

3. ERROR BEHAVIOUR OF PURE INERTIAL SYSTEMS

In a pure inertial system thrust acceleration represents the only input information measured at the first level. In most cases the input axes of the sensors are directed parallel to the axes of the navigational coordinate frame. If this is not the case, a coordinate transformation is necessary.

Figure 4 shows the block diagram of a pure inertial system used for navigation over the surface of the earth. In this case the velocity $\vec{v}$ of a vehicle is defined as the rate of change of the vector $\vec{R}$ with respect to the earth. ($\vec{v} = d\vec{R}/dt|_E$) . Therefore, instead of Equation (4), the following is now the relation for acceleration corrections:

$$\vec{b} = \left(\frac{d\vec{v}}{dt}\right)_N = \vec{a} + \vec{g}_m - \vec{\Omega} \times (\vec{\Omega} \times \vec{R}) - (2\vec{\Omega} + \vec{\rho}) \times \vec{v} , \qquad (5)$$

where $\vec{\Omega}$ = earth rate vector

$\vec{\rho} = \vec{\omega} - \vec{\Omega}$ = rate vector of the navigational coordinate frame with respect to an earth-fixed frame.

As can be seen from Figure 4, a feedback loop exists, owing to the gravity vector $\vec{g}$. i.e. the system is capable of oscillating. The oscillation occurs with the Schüler period

$$T_s = 2\pi\sqrt{\left(\frac{R_0}{g}\right)} .$$

The error behaviour of a pure inertial system in a Schüler-tuned mode is now discussed.

Figure 5 shows the error block diagram for one horizontal channel of such a system. The errors considered are accelerometer bias, gyro drift, and azimuth misalignment.

The following notation is used in Figure 5:

$\Delta\dot{x}$ = computed velocity error in x-direction

Δx = computed position error in x-direction

B_x = x-accelerometer bias

ϵ_y = fixed drift of the y-gyro

Φ_y = y-component of the error angle vector between the navigational and platform coordinate frame

ψ_z = z-component of the error angle vector between the platform and computer coordinate frame (azimuth misalignment).

The velocity and position errors, as derived from Figure 5, are

$$\Delta\dot{x} = \frac{sB_x - g(\epsilon_y + \omega_s\psi_z)}{s^2 + \omega_s^2} \qquad (6)$$

$$\Delta x = \frac{sB_x - g(\epsilon_y + \omega_s\psi_z)}{s(s^2 + \omega_s^2)} \qquad (7)$$

The characteristic equation (denominator of Equation (6)) is that of an undamped oscillation with Schüler frequency. For constant sensor errors, the mean value of the velocity error is

$$\Delta\dot{x}\Big|_{s=0} = -\frac{g}{\omega_s^2}(\epsilon_y + \omega_x\psi_z) \; . \tag{8}$$

The position error constantly increases with time,

$$\Delta x\Big|_{s=0} \rightarrow \infty \; . \tag{9}$$

The unbounded position error is due to gyro drift and azimuth misalignment; accelerometer bias does not cause an unbounded position error.

4. UPDATING METHODS AND RELATED ERROR BEHAVIOUR

Although Schüler-tuning is of great importance in navigation over the surface of the earth, it has been shown that for longer mission times the position errors can become intolerably large, especially because of gyro drift and azimuth error. To avoid this, the pure inertial system can be aided with additional or redundant navigation information from other sources. Continuous or discrete data available for the navigation computer may be of level 2, 3a and 3b. No additional information on level 1 is available, because we do not know any other method of measuring acceleration except by inertial means.

Accordingly, we distinguish different kinds of aided systems:

1. Velocity aided systems (redundant data at level 2).
2. Position aided systems (redundant data at level 3a).
3. Stellar monitored systems (redundant data at level 3b).
4. Stellar monitored and velocity damped systems (redundant data at level 2 and 3b).

The additional or redundant data can be of intermittent nature and used to replace the computed data at certain time periods. If it is always available, it can be continuously compared with the computed information; the difference signal, multiplied by various gains K_ν, can be added to the accelerometer output signals or the gyro torquing commands either directly or over additional integrators. If the difference or error signal is introduced at the next lower level, it is used to damp the oscillations of the system. If it is introduced at the same level inside the Schüler loop, then the frequency and amplitude can be changed. If additional integrators are used, sensor biases can be compensated.

Figure 6 shows the x-channel of a Schüler-tuned system and the mechanisation of the different updating methods.

These methods are now discussed in more detail, if Doppler velocity is available as redundant information. The Doppler velocity signal cannot very well be used to reset the inertial velocity signal directly, because of its high noise level. The direct resetting method is not very useful and is only feasible if angular measurements or position fixes are available.

As Figure 6 indicates, the difference signal can be used to provide three different correction signals, C_1, C_2 and C_3, into the Schüler loop.

The first signal C_1, the difference signal multiplied by the gain factor K_1, is added to the accelerometer output. This provides damping of the Schüler oscillation. The second signal C_2, the difference signal multiplied by the gain factor K_2/R, has an influence upon the oscillation frequency by adding it to the gyro input. The third input C_3 provides compensation for some steady-state errors of velocity by introducing an additional integrator with gain K_3/R.

Assuming that both sources of velocity information - the computed inertial and the measured reference velocity - are subjected to errors, the three corrections corresponding to the x-channel error block diagram can be written as follows:

$$C_{1x} = -K_1(v_{cx} - v_{rx}) = -K_1(\Delta\dot{x} - \delta v_x) \tag{10}$$

$$C_{2x} = \frac{K_2}{R}(v_{cx} - v_{rx}) = \frac{K_2}{R}(\Delta\dot{x} - \delta v_x) \tag{11}$$

$$C_{3x} = \frac{K_3}{R}\int(v_{cx} - v_{rx})\,dt = \frac{K_3}{R}\int(\Delta\dot{x} - \delta v_x)\,dt\,, \tag{12}$$

where $\Delta\dot{x}$ and δv_x represent the errors of the computed inertial and the measured Doppler reference velocity, respectively.

Figure 7 shows the x-channel error block diagram of the Doppler inertial system described.

The transfer functions of the velocity and position errors can be derived from Figure 7 as follows:

$$\Delta\dot{x} = \frac{s^2B_x - sg(\epsilon_y + \omega_x\psi_z) + [\omega_s^2(K_2s + K_3) + s^2K_1]\delta v_x}{s^3 + K_1s^2 + (1 + K_2)\omega_s^2s + K_3\omega_s^2} \tag{13}$$

$$\Delta x = \frac{s^2B_x - sg(\epsilon_y + \omega_x\psi_z) + [\omega_s^2(K_2s + K_3) + s^2K_1]\delta v_x}{s[s^3 + K_1s^2 + (1 + K_2)\omega_s^2s + K_3\omega_s^2]} \tag{14}$$

$$\Delta\dot{x}\big|_{s=0} = \delta v_x\,. \tag{15}$$

Equations (13) and (14) can be used to evaluate the velocity and position errors for different values of K_1, K_2 and K_3, or to obtain optimal values of these three gain factors. The steady-state velocity error now contains the reference velocity error δv_x only (Eq. (15)). The fixed gyro drift and azimuth error are compensated by the additional integrator.

Figure 8 shows, as an example, the time functions of the velocity error due to gyro drift for the pure inertial system $(K_1 = K_2 = K_3 = 0)$ and for the different methods of updating $(K_1 \neq 0;\ K_1$ and $K_2 \neq 0$ and $K_1, K_2, K_3 \neq 0)$.

So far constant errors (biases) have been discussed, but in many cases the errors are not constant, but rather stochastic processes characterised by their autocorrelation functions $g_\nu(\tau)$ and their spectral density functions $\Phi_\nu(\omega)$. For linear, non-time-varying systems, the statistics of the output signal $y(t)$ follow directly from the time-independent statistics of the input signal $x(t)$. The mean value of the output signal is zero if the mean value of the input signal is zero. The average value of the square of the output signal, the time-dependent variance $\sigma^2y(t)$, can be computed by the following expression[2]:

$$\sigma^2_{y(t)} = \int_0^t\!\!\int_0^t g(\tau_1)g(\tau_2)\varphi_x(|\tau_1 - \tau_2|)\,d\tau_1 d\tau_2\,. \tag{16}$$

where $g(\tau) = \frac{1}{2\pi j}\int_{c-j\infty}^{c+j\infty} G(s)e^{s\tau}ds$ = system impulse response

$G(s)$ = system transfer function

$\rho_x(|\tau|)$ = input signal autocorrelation function.

If the input signal is white noise, the autocorrelation function is given by

$$\rho_x(|\tau_1-\tau_2|) = A\delta(|\tau_1-\tau_2|) . \tag{17}$$

Equation (17) substituted into Equation (16) yields

$$\sigma^2_{y(t)} = \int_0^t\int_0^t g(\tau_1)g(\tau_2)A\delta(|\tau_1-\tau_2|)\, d\tau_1 d\tau_2$$

$$= A\int_0^t g(\tau_1)\, d\tau_1\int_0^t g(\tau_2)\delta(|\tau_1-\tau_2|)\, d\tau_2 .$$

The inner integral can be evaluated by using the distribution property of the δ-function

$$\int_0^t g(\tau_2)\delta(|\tau_1-\tau_2|)\, d\tau_2 = \begin{Bmatrix} g(\tau_1) & \text{for } \tau_1 \text{ in } [0,t] \\ 0 & \text{for } \tau_1 \text{ not in } [0,t] \end{Bmatrix}. \tag{18}$$

Using Equation (18), the expression for $\sigma^2_{y(t)}$ becomes

$$\sigma^2_{y(t)} = A\int_0^t g^2(\tau_1)\, d\tau_1 . \tag{19}$$

If the input signal is bandwidth limited white noise with autocorrelation function

$$\rho_x(|\tau|) = \sigma_x^2\, e^{-|\tau|/\tau_x} . \tag{20}$$

Equation (19) can also be used if τ_x is small compared with the system time constants. Equation (20) can then be approximated by

$$\rho_x(|\tau|) \simeq 2\tau_x\sigma_x^2\delta(\tau) \tag{21}$$

and the output variance yields, in this case,

$$\sigma^2_{y(t)} = 2\tau_x\sigma_x^2\int_0^t g^2(\tau)\, d\tau . \tag{22}$$

Another method of determining the steady-state value of the variance $\sigma^2_{y(t)}$, if it is convergent, is given by the following expression[2]:

$$\lim_{t\to+\infty} \sigma^2_{y(t)} = \frac{1}{2\pi}\int_{-\infty}^{+\infty} \Phi_y(\omega)\, d\omega . \tag{23}$$

The output spectral density function is given by

$$\Phi_y(\omega) = |G(j\omega)|^2\, \Phi_x(\omega) . \tag{24}$$

For bandwidth limited white noise, Φ_x can be evaluated from the approximated auto-correlation function $\rho_x(|\tau|)$ of Equation (21):

$$\Phi_x(\omega) = \int_{-\infty}^{+\infty} \rho_x(|\tau|)e^{-j\omega\tau}\, d\tau = 2\tau_x\sigma_x^2 \tag{25}$$

Equations (24) and (25), when substituted into Equation (23), yield

$$\operatorname*{Limit}_{t\to+\infty} \sigma^2_{y(t)} = (2\tau_x\sigma_x^2)\,\frac{1}{2\pi}\int_{-\infty}^{+\infty} |G(j\omega)|^2\, d\omega\ . \tag{26}$$

It should be noted that Equations (23) and (26) can only be used if $G(s)$ has no poles on the imaginary axis or in the right half-plane. A pure inertial system in a Schüler-tuned mode is undamped, i.e. it has poles on the imaginary axis. This means that the effect of noise inputs must be calculated according to Equations (16) or (22).

As an example, the variance of the velocity and position errors due to a white noise Doppler input error δv_x for a damped system are now calculated.

The transfer functions for $K_2 = K_3 = 0$ are

$$\sigma_{\Delta\dot{x}}(s) = \frac{K_1 s}{s^2 + K_1 s + \omega_s^2} \tag{27}$$

$$G_{\Delta x}(s) = \frac{K_1}{s^2 + K_1 s + \omega_s^2}\ . \tag{28}$$

First, Equation (26) is used:

$$\begin{aligned}
\sigma^2_{\Delta\dot{x}} &= 2\tau_{\delta v_x}\,\sigma^2_{\delta v_x}\,\frac{1}{2\pi}\int_{-\infty}^{+\infty} |G_{\Delta\dot{x}}(j\omega)|^2\, d\omega \\
&= 2\tau_{\delta v_x}\,\sigma^2_{\delta v_x}\,\frac{K_1^2}{\pi}\int_0^{\infty} \frac{\omega^2}{(\omega_s^2-\omega^2)^2 + K_1^2\omega^2}\, d\omega \\
&= 2\tau_{\delta v_x}\,\sigma^2_{\delta v_x}\,\frac{K_1\omega_s}{\sqrt{(4\omega_s^2-K_1^2)}}\cos\tfrac{1}{2}\arctan\frac{K_1\sqrt{(4\omega_s^2-K_1^2)}}{2\omega_s^2-K_1^2} \\
&= \tau_{\delta v_x}\,\sigma^2_{\delta v_x}\,K_1
\end{aligned} \tag{29}$$

$$\begin{aligned}
\sigma^2_{\Delta x} &= 2\tau_{\delta v_x}\,\sigma^2_{\delta v_x}\,\frac{K_1^2}{\pi}\int_0^{\infty} \frac{d\omega}{(\omega_s^2-\omega^2)^2 + K_1^2\omega^2} \\
&= \frac{\tau_{\delta v_x}\,\sigma^2_{\delta v_x}\,K_1}{\omega_s^2}\ .
\end{aligned} \tag{30}$$

Now Equation (22) is used to determine the time-dependent variances.

The system impulse responses become

$$g_{\Delta\dot{x}}(t) = K_1 e^{-(K_1/2)t} \cos \omega_0 t - \frac{K_1}{2\omega_0} \sin \omega_0 t = \dot{g}_{\Delta x}(t) \tag{31}$$

$$g_{\Delta x}(t) = \frac{K_1}{\omega_0} e^{-(K_1/2)t} \sin \omega_0 t \, , \tag{32}$$

where $\omega_0 = \frac{1}{2} \sqrt{(4\omega_s^2 - K_1^2)}$. (33)

The desired time-dependent variances are obtained as follows:

$$\sigma^2_{\Delta\dot{x}}(t) = 2\tau_{\delta v_x} \sigma^2_{\delta v_x} \int_0^t g^2_{\Delta\dot{x}}(\tau)\, d\tau$$

$$\sigma^2_{\Delta\dot{x}}(t) = \frac{2\tau_{\delta v_x} \sigma^2_{\delta v_x} K_1^2}{\omega_0(4\omega_0^2 + K_1^2)} \left[\frac{\omega_0(4\omega_0^2 + K_1^2)}{2K_1} - e^{-K_1 t} \left(\frac{8\omega_0^2 K_1^2 + 16\omega_0^4 + K_1^4}{8\omega_0 K_1} , \right. \right.$$

$$\left. \left. - \frac{4\omega_0^2 + K_1^2}{4} \sin 2\omega_0 t - \frac{K_1^4 + 4\omega_0^2 K_1^2}{8\omega_0 K_1} \cos 2\omega_0 t \right) \right] \tag{34}$$

$$\sigma^2_{\Delta x}(t) = 2\tau_{\delta v_x} \sigma^2_{\delta v_x} \int_0^t g^2_{\Delta x}(\tau)\, d\tau$$

$$\sigma^2_{\Delta x}(t) = \frac{2\tau_{\delta v_x} \sigma^2_{\delta v_x} K_1^2}{\omega_0(4\omega_0^2 + K_1^2)} \left[\frac{2\omega_0}{K_1} - e^{-K_1 t} \left(\frac{4\omega_0^2 + K_1^2}{2\omega_0 K_1} + \sin 2\omega_0 t - \right. \right.$$

$$\left. \left. - \frac{K_1}{2\omega_0} \cos 2\omega_0 t \right) \right] . \tag{35}$$

If K_1 is small compared with ω_s , the following approximations for $\sigma^2_{\Delta\dot{x}}(t)$ and $\sigma^2_{\Delta x}(t)$ result:

$$\sigma^2_{\Delta\dot{x}}(t) \simeq \tau_{\delta v_x} \sigma^2_{\delta v_x} K_1 (1 - e^{-K_1 t}) \tag{36}$$

$$\sigma^2_{\Delta x}(t) \simeq \frac{\tau_{\delta v_x} \sigma^2_{\delta v_x} K_1}{\omega_s^2} (1 - e^{-K_1 t}) \, . \tag{37}$$

Equations (36) and (37) show that the "filtering" of the Doppler noise through the inertial system depends largely on the gain factor K_1 . Also, the time constant for the variance of the velocity and the position error are inversely proportional to K_1 . These facts give additional criteria for K_1 . If K_2 and K_3 are also $\neq 0$, it is difficult to obtain a closed-form solution of Equation (22). In this case a computer programme may be necessary to evaluate the integral and to obtain criteria for the values of K_2 and K_3 .

5. LINEAR OPTIMAL FILTERING OF REDUNDANT DATA

The updating mechanism described in Section 4 used a deterministic approach. The advent of very powerful navigation computers made it possible to retrieve more of the information contained in the navigation measurements by applying statistical filtering techniques.

In this section the optimum estimation problem of the state variables of aided inertial systems is discussed. For the optimum estimation a so-called Kalman or minimum variance filter is used to derive a "best" estimate of the system state, that is, optimum use is made of all available information in the least-square sense. The properties of such filters can be summarised as follows:

The filter estimates the errors of the system state variables.

The estimate is based upon statistical data of error measurements and state variables.

The filter formulae satisfy minimum variance criteria for all problem parameters.

The formulae implemented are recursive. This means that the optimum estimate for the time being can be computed from the previous estimate and the current observation without recourse to earlier estimates or observations.

The recursive formulae are well suitable for computation on digital computers.

The basic method of using a linear optimum estimation filter is shown in Figure 9.

The two upper blocks in Figure 9 represent a pure inertial system. The navigation equations are mechanised in the navigation computer in the conventional way. Compensation for sensor bias may be included in the navigation equations.

The information available in the navigation computer is compared with redundant external data. The difference is the measurement data used in the filter to improve the estimate of the errors of the system.

These error estimates of the filter may be used either for updating of the navigation information and thus for correction of Inertial Measuring Unit (IMU) errors (closed-loop operation), or for improvement of the error model (open-loop operation) or both. The block "controller" regulates this operation.

Because some of the errors of an inertial system are unbounded, closed-loop operation is necessary in all cases where inertial systems are used for longer mission duration.

The derivation of the Kalman filter equations is omitted, because it is given in the literature$^{3-5}$. Their basic elements are now discussed. The system can be described by a set of linear, first-order differential equations, which may be written in matrix notation as follows:

$$\dot{X} = A(t)X(t) + N(t) \tag{38}$$

$$Y = M(t)X + v(t) \,. \tag{39}$$

Equation (38) describes the error propagation of an inertial navigator with $N(t)$ being the forcing function, representing a white noise process. This set of equations can be obtained by expansion of the error equations about the indicated position, velocity and orientation of the navigator. Equation (39) is the measurement equation. It relates the values of the error state vector to the measurement data. $v(t)$ represents the measurement noise.

Since continuous equations are not suitable for computation on digital computers, it is necessary to transform Equations (38 and (39) into the discrete form

$$X_{K+1} = \Phi_{K+1,K} X_K + h_K \tag{40}$$

$$Y_K = M_K X_K + v_K \tag{41}$$

where X_K = system error state vector at time t_K, containing velocity, position and platform angle errors, error contributing parameters like gyro drifts, misalignment angles, accelerometer bias, and measurement variables like Doppler bias etc. The state vector usually consists of 7 or 9 navigation states, 7 or more IMU states, and 2 or more measurement states.

$\Phi_{K+1,K}$ = state transition matrix relating X_{K+1} to X_K in the noise-free case.

h_K = additive white noise sequence in the dynamical system at time t_K.

Y_K = observation vector of measurement data at time t_K linearly related to X_K.

M_K = measurement matrix at time t_K relating Y_K to X_K in the noise-free case.

v_K = additive white noise sequence corrupting measurement data at time t_K.

The following five equations represent the basic filter:

$$\bar{X}_{K+1} = \Phi_{K+1,K} \hat{X}_K \tag{42}$$

$$\hat{X}_K = \bar{X}_K + B_K(Y_K - \bar{Y}_K) \tag{43}$$

$$B_K = P_K^* M_K^T [M_K P_K^* M_K^T + V_K]^{-1} \tag{44}$$

$$P_{K+1}^* = \Phi_{K+1,K} P_K \Phi_{K+1,K}^T + H_K \tag{45}$$

$$P_K = (I - B_K M_K) P_K^* \quad , \quad \text{if } B_K \text{ of Equation (44) is used ,} \tag{46}$$

where $\bar{Y}_K$ = $M_K \bar{X}_K$.

$\bar{X}_K$ = *a priori* estimate of the system state predicted at time t_K before using measured data.

$\hat{X}_K$ = *a posteriori* estimate of the system state at time t_K after using measured data.

P_K = covariance matrix of the error state vector $X - \hat{X}_K$.

P_K^* = covariance matrix of the error state vector $X - \bar{X}_K$.

H_K = $E[h_K h_K^T]$ = covariance matrix of the system noise.

V_K = $E[v_K v_K^T]$ = covariance matrix of measurement noise.

B_K = weighting matrix at time t_K .

These equations are processed on an iterative basis to determine consecutively the optimum estimate $\hat{X}_K$ of the system state at time t_K .

Figure 10 is a block diagram of the Kalman filter computations occurring during the period t_{K-1} to t_{K+1}. Block ① is a mathematical model of the system described by Equations (40) and (41). The state of this system is to be estimated by block ②. In this estimation block, the measurements Y_k corresponding to the *a priori* estimate, $\bar{X}_K$, are compared with the real measurements Y_K. The difference multiplied by the weighting matrix B_K is then used to determine the desired *a posteriori* estimate $\hat{X}_K$, as given by Equations (42) and (43). The weighting matrix computed in block ③ depends on the statistical data of the plant noise h_K and the measurement noise v_K. The error covariance matrices P_K and P_K^* are also computed in block ③ by the use of the recursive formulae (44), (45) and (46).

Some remarks are now made in conjunction with the problems of filter mechanisation and initial conditions. It is not possible to cover this area exhaustively.

The system state vector should include all the system variables to be estimated. Since the size of computation grows excessively (approximately by the third power) with the number of variables, appropriate choice is of great importance. In very accurate or complicated systems, the number of the state variables may be too large to be processed on the computer used. In such cases, a sub-optimal system may be used, where some of the variables are omitted. Another method is the partitioning of the system state vector by placing the variables to be eliminated in one sub-system and the remaining variables in an additional sub-system[6]. In every case of designing a sub-optimal filter, the performance deterioration due to elimination or partitioning of some state variables should be predetermined by digital computer simulation and compared with optimal results.

The state transition matrix Φ_{t,t_0} is derived from the state matrix $A(t)$ by the differential equation

$$\frac{d\Phi_{t,t_0}}{dt} = A(t)\Phi_{t,t_0} , \tag{47}$$

with the initial condition $\Phi_{t_0,t_0} = I$ (the identity matrix). An explicit solution of this equation is not possible.

However, if $A(t)$ is slowly varying, that means all coefficients can be assumed constant within one computation cycle, various expressions for Φ can be obtained. In particular the solution for Φ in the time interval t_1 to t_2 for constant $A = A(t_1)$ is

$$\Phi_{t_2,t_1} = e^{(t_2-t_1)A(t_1)} . \tag{48}$$

If the time interval t_2-t_1 is very small compared to the system dynamics, then

$$\Phi_{t_2,t_1} = I + (t_2-t_1)A(t_1) , \tag{49}$$

where higher order elements are neglected.

If the Kalman computation cycle $\Delta t = t_{K+1} - t_K$ is too large to satisfy the condition (49), it can be partitioned by computing Φ_{t_{i+1},t_i} over smaller time intervals $\Delta t_i = t_{i+1} - t_i$, where $\Delta t = p\Delta t_i$. Since

$$\Phi(t_2,t_0) = \Phi(t_2,t_1)\Phi(t_1,t_0) ,$$

we obtain, for $\Phi_{K+1,K}$,

$$\Phi_{K+1,K} = \prod_{i=1}^{p} \Phi_{i,i-1} \tag{50}$$

The driving functions in Equations (40) and (41) are restricted to white noise. If some of the state variables are affected by correlated noise, this can also be included by the addition of shaping filters. These filters are then considered to be excited by white noise and their outputs have the same correlations as the correlated random variables in the original system.

As a simple example, Markov noise processes have a spectral density function of the form

$$G(\omega) = \frac{2\sigma^2\omega_0}{\omega^2 + \omega_0^2} \tag{51}$$

and can be described by the differential equation

$$\dot{f}(t) + \omega_0 f(t) = \sqrt{(2\sigma^2\omega_0)}\,u(t) \, , \tag{52}$$

where $u(t)$ is unity white noise. This equation can be incorporated in Equation (38), and $f(t)$ is one of the state variables. It may be emphasised that extension to correlated noise is not compromised by the increase in the number of state variables.

For the initial conditions, the variances and covariances to be arrayed in the covariance matrix P_0^* should be known. Since, in many cases, the cross-correlations will not be known, the initial system may be obtained by replacing the symmetric matrix P^* by a diagonal matrix with only self-correlations, as in Equation (53):

Covariance Matrix of $\bar{X}$

$$P_0^* = \begin{bmatrix} \sigma_{x1}^2 & 0 & . & . & . \\ 0 & \sigma_{x2}^2 & . & . & . \\ 0 & 0 & \sigma_{x3}^2 & . & . \\ . & . & . & . & . \\ . & . & . & . & \sigma_{xn}^2 \end{bmatrix} . \tag{53}$$

During the Kalman filtering operation, the P_k and P_k^* matrices will converge to the optimal case as a result of the updating.

It has already been proved, by extensive system simulations and actual tests, that the optimum filtering method is a powerful tool to make maximum use of all available sensor information. But there is no theoretical or closed-form solution to find the "best" optimum filter for a particular system. Extensive computer simulations and, if possible, actual hardware tests are required to optimise the filter parameters and to determine the number of variables that should be estimated.

REFERENCES

1. Pitman, G.R.	*Inertial Guidance.* John Wiley, New York, London, 1962.
2. Davenport, W.B. Root, W.L.	*Random Signals and Noise.* McGraw-Hill, New York, 1958.
3. Kalman, R.E.	*A New Approach to Linear Filtering and Prediction Problems.* Journal of Basic Engineering, Vol.82, No.1, March 1960.
4. Sorenson, H.W.	*Kalman Filtering Techniques.* Chapter 5 of "Advances in Control Systems", Vol.3, edited by C.T.Leondes, Academic Press, New York, 1966.
5. Fagin, S.L.	*Recursive Linear Regression Theory, Optimal Filter Theory, and Error Analysis of Optimal Systems.* Institute of Electrical and Electronic Engineers, International Convention Record, Vol.12, Part 1, New York, March 1964, pp.216-245.
6. Pentcost, E.E. Stubberud, A.R.	*Synthesis of Computationally Efficient Sequential Linear Estimators.* Institute of Electrical and Electronic Engineers, Transactions, Vol.AES-3, No.2, March 1967.

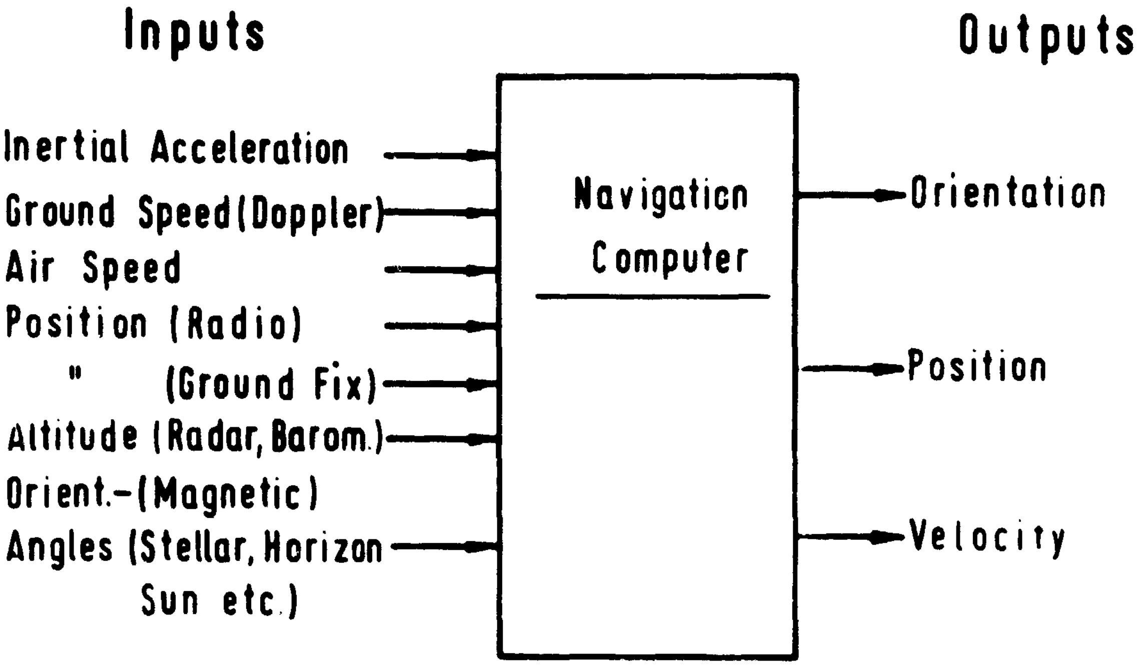


Fig.1 Inputs and outputs of navigation systems

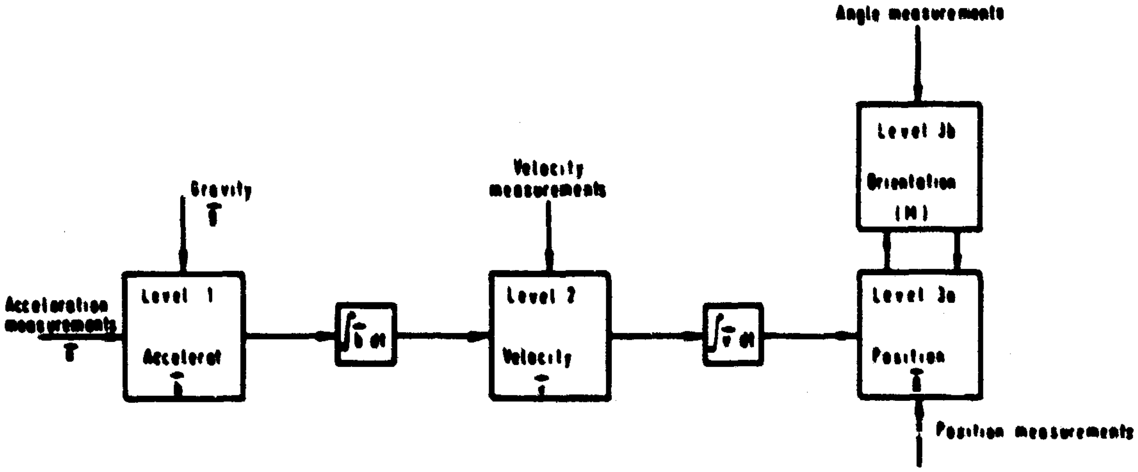


Fig.2 Levels of navigation information

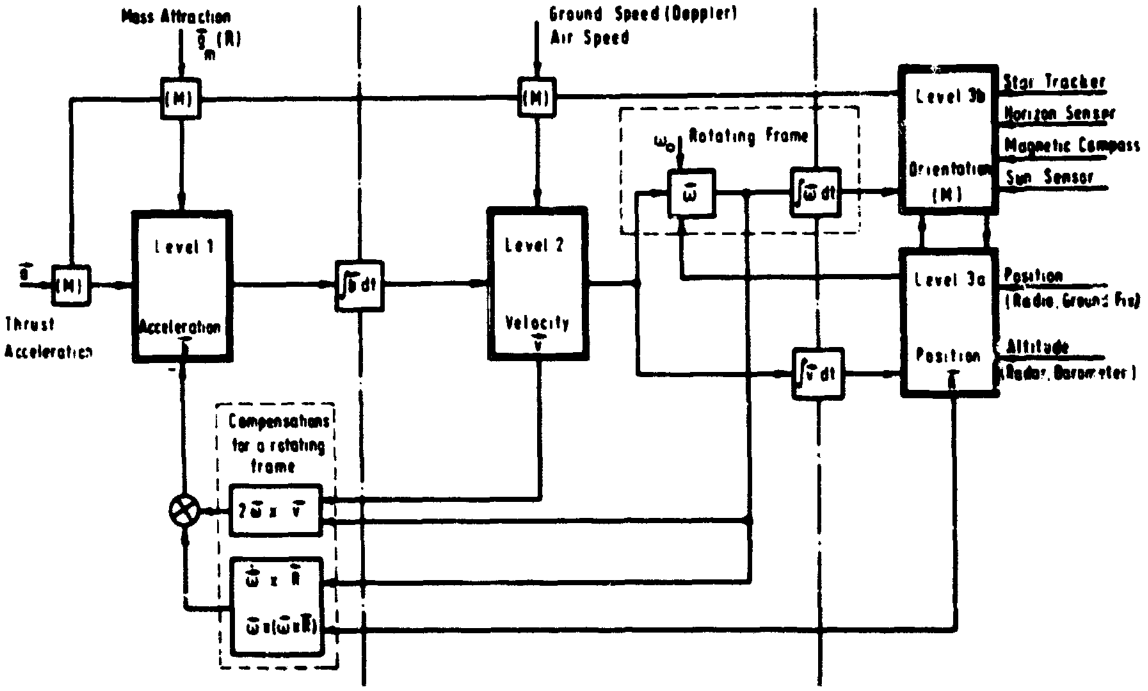


Fig.3 General block diagram of navigation systems

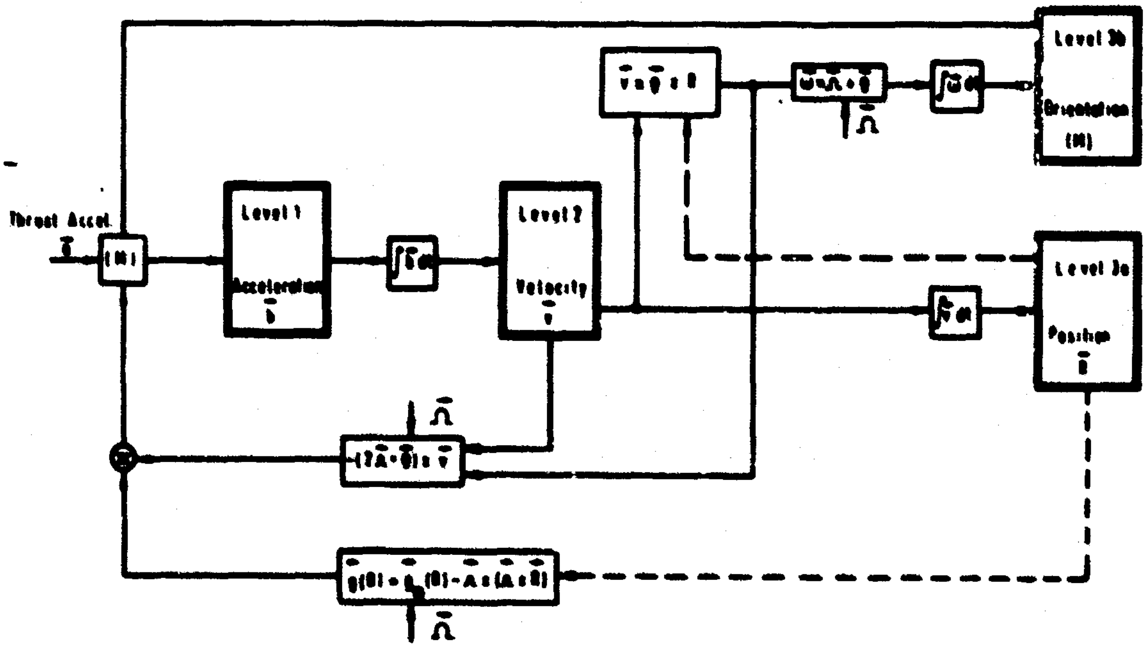


Fig.4 Vector block diagram of a pure inertial system used for earth navigation

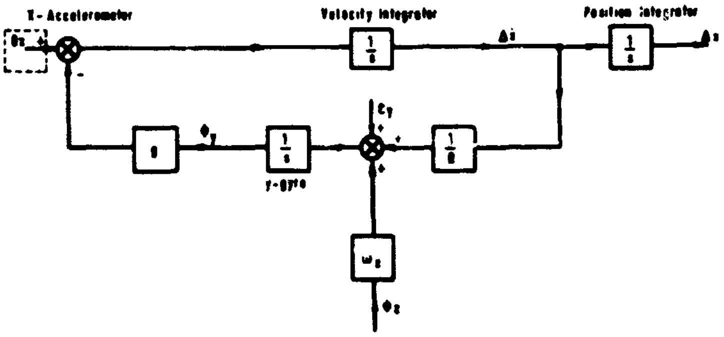


Fig. 5 Error block diagram for the x-channel of a Schüler-tuned system

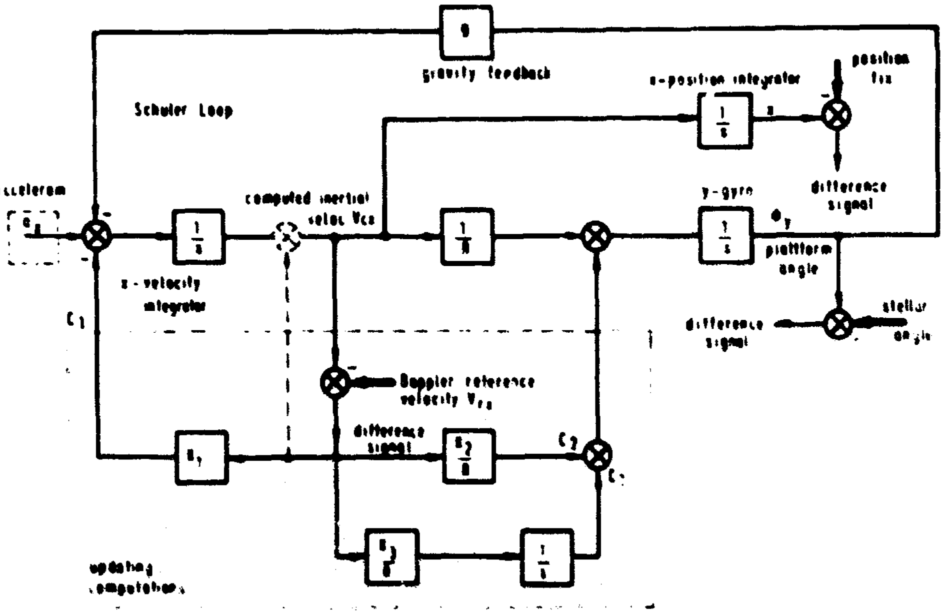


Fig. 6 Updating mechanisation

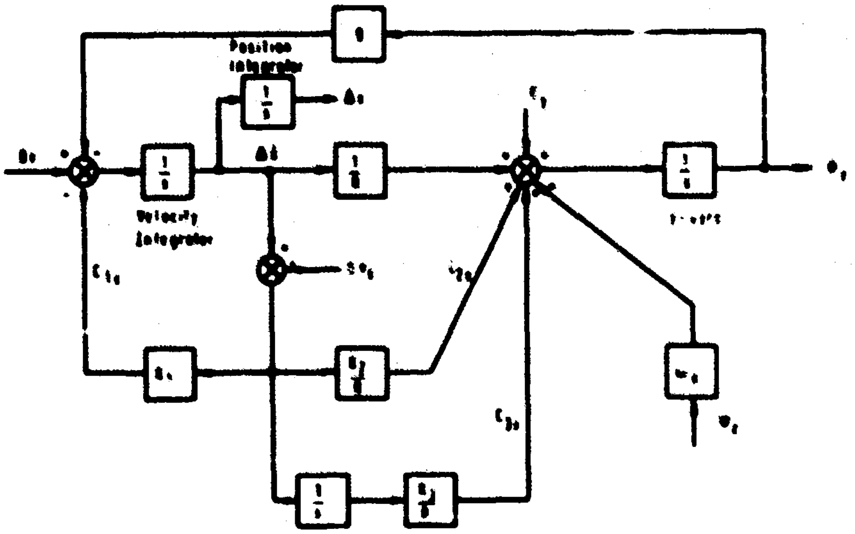


Fig. 7 Error block diagram of a Doppler inertial system (x-channel)

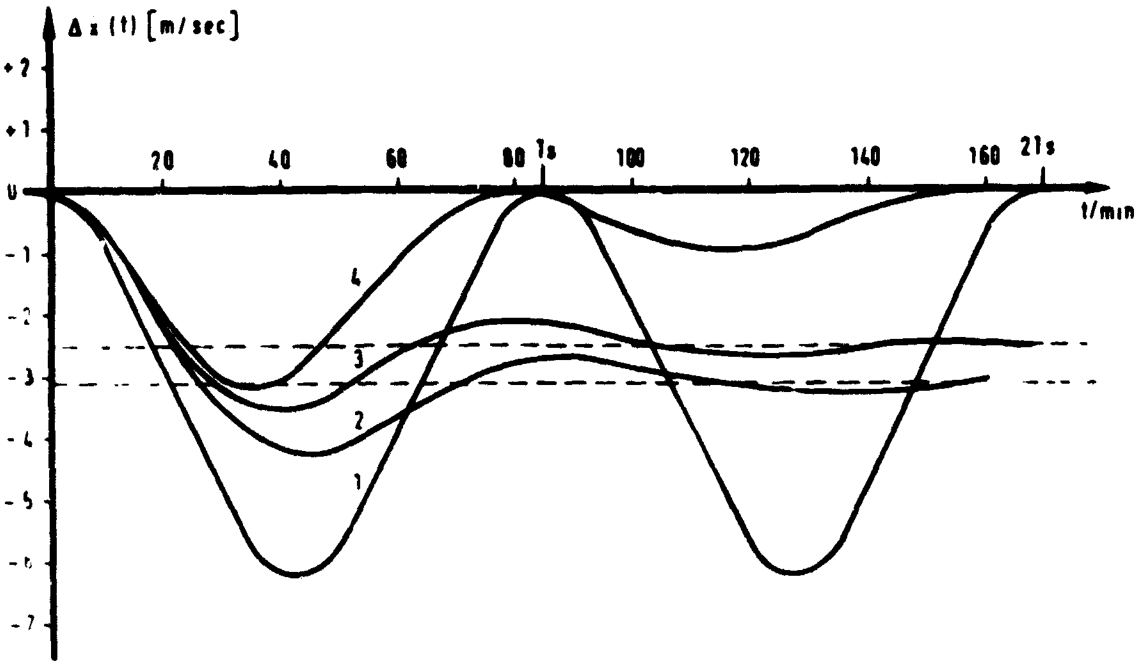


Fig.8 Velocity error due to fixed gyro drift ($\epsilon_y = 0.1°/h$) for different systems

1: Pure inertial system ($K_1 = 0$; $K_2 = 0$; $K_3 = 0$)

2: Doppler damped inertial system ($K_1 = 0.6\omega_s$; $K_2 = 0$; $K_3 = 0$)

3: Doppler inertial system ($K_1 = 0.6\omega_s$; $K_2 = 0.236$; $K_3 = 0$)

4: Doppler inertial system with an additional integrator ($K_1 = 0.6\omega_s$; $K_2 = 0.236$; $K_3 = 0.2312\omega_s$)

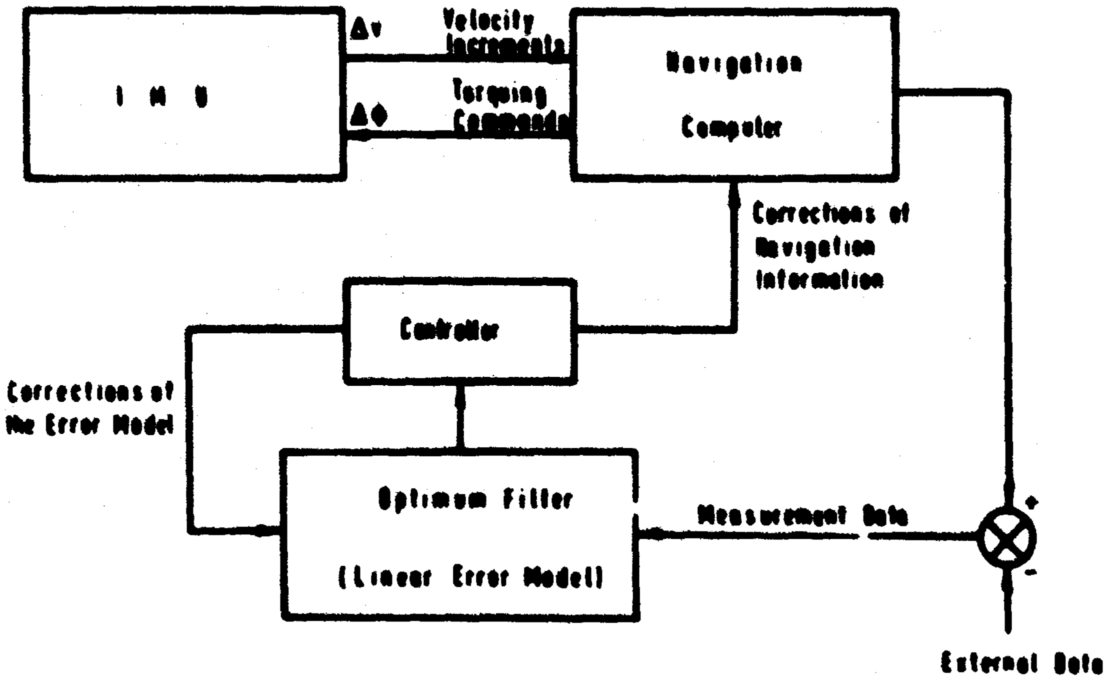


Fig.9 Block diagram of an inertial system with optimum filter

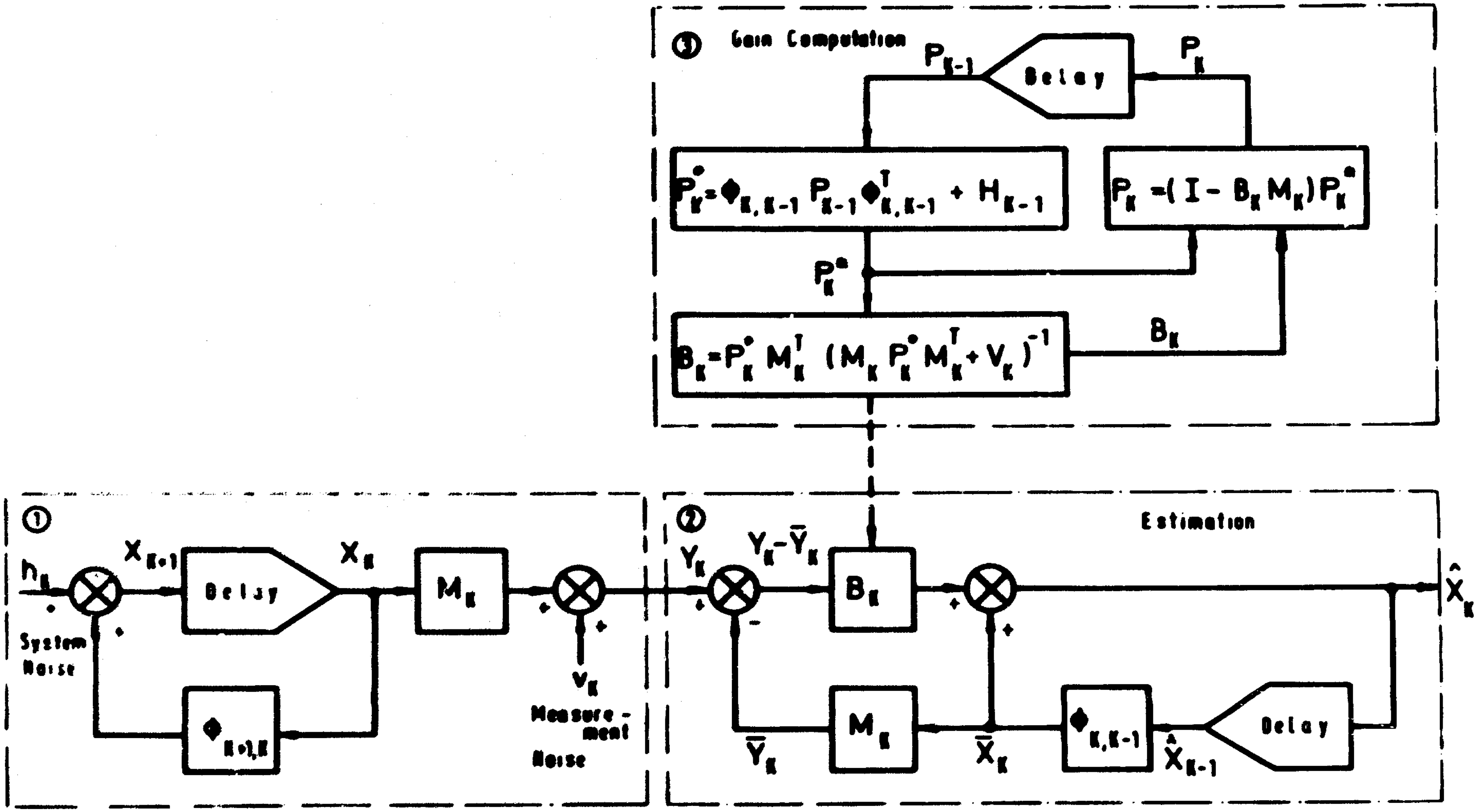


Fig.10 Block diagram of linear optimum filter

THE IMPACT OF STATISTICAL ESTIMATION ON INERTIAL NAVIGATION

by

Larry D. Brock

Air Force Missile Development Center,
Holloman Air Force Base, New Mexico, USA

SUMMARY

The application of statistical estimation techniques is one of the more important new developments in inertial navigation. These techniques make possible significant improvement in the performance of a navigation system, even with no change in the quality of the instruments, by allowing the system to continuously calibrate itself. The purpose of this paper is to describe the operation of an onboard statistical filter and discuss some of the practical consequences and problems.

First, the basic form of the Kalman recursive filter is reviewed. The basic operation of the filter is then explained using a simplified example. Next, simulation results are shown for a typical air transport navigation problem with an inertial system aided by Doppler radar and position fixes. Several applications and consequences of statistical estimation are then described, and two of the more important practical problems in applying statistical estimation procedures are discussed.

THE IMPACT OF STATISTICAL ESTIMATION ON INERTIAL NAVIGATION

Larry D.Brock

1. INTRODUCTION

One of the most important new developments in inertial navigation systems is the use of statistical estimation techniques during flight. These techniques are programmed into a flight computer and provide for the optimum use of all navigation information. The computer program is normally called a filter because of its close association with a conventional filter circuit. The computer program separates the navigation information (signal) from the errors in the navigation equipment (noise). The statistical filtering technique in most common use, which was developed by R.E.Kalman[1,2], is closely related to the statistical filter work of N.Wiener[3].

The primary applications of these techniques are in the incorporation of external information, such as Doppler velocity and position fixes, into an inertial system and in the initial alignment and calibration of the inertial platform. Statistical filtering will, in most cases, give significant improvement in the performance of a navigation system, even with no change in the quality of the instruments. This improvement is made possible through the use of a computer technique which not only combines the output of the various systems (inertial, Doppler, etc.) in an optimum way, but also provides a means by which the systems continuously calibrate each other. The purpose of this paper is to describe the operation of an onboard statistical filter and to discuss some of the practical consequences and problems resulting from its use.

First, a brief intuitive explanation is given of how the filter works; then the equations for the Kalman formulation of a statistical filter are summarized. The filter is further explained by the application of the equations to a simple example. Next, simulation results are given for a typical New York to London flight using an inertial system with Doppler radar and position fixes. Finally, some of the more significant results, consequences, and problems are discussed.

2. DESCRIPTION OF STATISTICAL FILTER

As it is commonly used in a navigation system a statistical filter is a recursive technique programmed into a flight computer which uses all available information to estimate the errors and error sources in the system. At each time new information is obtained, a new estimate is made of the errors and error sources. The estimated error is subtracted from the output of the system and the estimated error sources (gyro drift, accelerometer scale factor, etc.) are compensated for. The information used by the filter is the difference between the basic navigation outputs as produced by the different systems. Examples are the difference between the velocity as measured by the inertial system and that indicated by the Doppler radar or the difference between the position as indicated by the inertial system and that indicated by LORAN. The essential ingredient of the filter is a statistical error model for all the systems involved (inertial, Doppler, etc.) The computer contains a current estimate for each term in the error model. This estimate, based on all previous measurements, is periodically updated. At the time of each new measurement, the difference in the outputs of the systems is predicted, based on the current estimate of the errors in the systems. If there is a difference between the predicted difference in the systems and the measured difference, then the estimates of the errors are not as good as they could be. This difference between the predicted and actual measurements is used through a set of weighting coefficients to update each of the estimates of the errors. The weighting coefficients are variables which are computed periodically in the flight computer and are based on the assumed statistical model for the errors. This computation takes into account the past history of the system, including the effects of previously applied information and of vehicle motions which affect the system errors.

The basic operation of a filter can be illustrated by the following hypothetical situation: Suppose a human navigator is monitoring the operation of an inertial system on a long flight. Also, suppose that periodically he gets position fixes from other sources -- such as LORAN, TACAN, or celestial sightings -- and plots the difference between the output of the inertial system and the fixes. Now if the human navigator has had considerable experience with inertial systems, he will recognize that the relatively smooth trend in this difference plot (usually with an 84-minute period) is primarily the inertial system error and that the random errors are primarily the fix uncertainties. He can also probably tell by the shape of the inertial error curve what the probable causes of the errors are. For example, he will know that if there is an azimuth misalignment, the error curve will change in a certain way just after a turn. The navigator can then compensate for the errors. It would be very impractical for even a highly trained human operator to do this much work during an actual flight, but this is essentially what the filter does automatically.

3. THE KALMAN STATISTICAL FILTER

The entire process discussed in the preceding section can be described mathematically with a relatively concise set of equations developed by R.E.Kalman[1,2]. These equations specify the optimum statistical filter if certain assumptions can be made about the system. These assumptions are that the uncertainty in the estimates of the system variables can be described by a set of linear equations that are driven by uncorrelated (white) noise. These assumptions imply that the uncertainties can be described by Gaussian distributions. The assumptions are not very restrictive and in almost all navigation problems they can be made with little loss in the efficiency of the filter.

The Kalman filter has become relatively familiar; thus no detailed derivation is given here. The equations for the Kalman filter are given as a reference and to define notation.

The basic filtering problem is to produce the optimum estimate of the fundamental variables (state variables) of a physical system from a set of measurements. The entire system is assumed to be described by equations of the form:

$$\dot{\vec{x}} = F\vec{x} + \vec{n} \tag{1}$$

where the dot indicates differentiation with respect to time. The vector $\vec{x}$ is the state vector and includes all the variables of all the systems involved in the measurement process. The matrix F is made up of the coefficients of the linear set of differential equations. (It is not necessary that the basic equations be linear. If they are not linear, the matrix F is then made up of the partial derivatives of the basic equations with respect to changes in the state variables about a nominal. For this discussion it is assumed that the basic equations are linear.) The vector $\vec{n}$ represents the white noise which drives the system. True white noise does not exist in the physical world since it implies infinite power. The white noise used in Equation (1) is the input into an analytical system (shaping filter) which produces the noise that is encountered in the real system. For example, if it is assumed that gyro drift rate is a stationary random process with an exponential autocorrelation function, then the shaping filter would be as shown in Figure 1, where w is white noise, d_r is the random drift rate, and τ is the correlation time of the drift. If it is assumed that gyro drift rate is described by a random walk (Brownian motion) process, then the shaping filter would be just the integral of white noise without the feedback. In order to fit the total system into the framework for a Kalman filter, it is necessary to include the basic variables of all the shaping filters into the state vector $\vec{x}$.

It is assumed that measurements are made at discrete times and are given by

$$\vec{m} = H\vec{x} + \vec{u} , \tag{2}$$

where the matrix H gives the relationship between the measurements and the state variables and $\vec{u}$ represents uncorrelated errors in the measurements. The optimum linear filter developed by Kalman for obtaining the optimum estimate, $\hat{\vec{x}}$, of the state variables $\vec{x}$, from a series of measurements, $\vec{m}$, is given by the set of equations:

$$\left.\begin{array}{l}\left.\begin{array}{rcl}\hat{\vec{x}} & = & \hat{\vec{x}}' + W(\vec{m} - H\hat{\vec{x}}') \\ E & = & E' - W HE' \\ W & = & E'H^T (H E' H^T + R)^{-1}\end{array}\right\} \text{at measurement times} \\ \left.\begin{array}{rcl}\dot{\hat{\vec{x}}} & = & F\hat{\vec{x}} \\ \dot{E} & = & FE + EF^T + Q ,\end{array}\right\} \text{between measurements}\end{array}\right\} \tag{3}$$

In these equations the prime indicates conditions that exist just prior to the measurement. The superscript T indicates the transpose. The matrix E is the covariance matrix of the estimation errors and is defined by

$$E = \overline{\delta x \; \delta x^T}$$

where

$$\delta\vec{x} = \hat{\vec{x}} - \vec{x} ,$$

The bar represents the expected or mean value. The matrices R and Q are defined by

$$\overline{u(t)\; u(\tau)} = R\, \delta(t - \tau)$$

$$\overline{n(t)\; n(\tau)} = Q\, \delta(t - \tau) ,$$

where δ is the unit impulse function (delta function). A diagram of a system with a Kalman filter is shown in Figure 2. A system with a Kalman filter can be mechanized in two almost equivalent ways. One way is to compensate for the estimated errors by simulating their effect in the flight computer and correcting the displays accordingly. The other method is to actually correct the error source in the physical system. For example, if an estimate is made of a gyro drift, then the gyro is torqued to remove this drift. Mathematically the two methods are equivalent, but in an actual system there may be some advantage in using one or the other.

4. EXAMPLE

A simple example will help illustrate the basic operation of a statistical filter. Suppose an inertial navigation system is built to navigate only along a meridian of a spherical non-rotating earth. The only sources of error are assumed to be an initial tilt error and gyro drift. The gyro drift has an initial bias value and varies randomly during the flight as a random walk (Brownian motion). The inertial system is to be aided by a Doppler radar with a scale factor error and uncorrelated errors at sample times. The scale factor error is also described by random walk. The inertial velocity and Doppler velocity are to be compared and the difference is to be used in a Kalman filter to produce corrections to the inertial and Doppler systems. A block diagram showing the operation of the system is given in Figure 3.

In the figure v_d is the Doppler velocity, ϵ_v is the uncorrelated error in the Doppler measurement, a is measured acceleration, v_i is inertial velocity, p is position, and R_e is the radius of the earth. The small circled numbers indicate where corrections from the filter go back into the system.

The equations which describe the errors in the inertial systems are

$$\left.\begin{aligned} \Delta\dot{p} &= \Delta v_i \\ \Delta\dot{v}_i &= g\Delta\alpha \\ \Delta\dot{\alpha} &= -\frac{\Delta v_i}{R_e} + \Delta d \\ \Delta\dot{d} &= n_d \, , \end{aligned}\right\} \qquad (4)$$

where $\Delta\alpha$ is the tilt of the platform relative to the vertical, g is the acceleration of gravity, Δd is the uncompensated gyro drift, and n_d is the white noise which gives gyro drift the proper random walk characteristics. The error in the Doppler velocity is

$$\left.\begin{aligned} \Delta v_d &= v\,\Delta SF + \epsilon_v \\ \Delta\dot{SF} &= n_s \, , \end{aligned}\right\} \qquad (5)$$

where v is the total velocity, ΔSF is the scale factor error, and n_s is white noise. The basic measurement is the difference in velocity,

$$\Delta v = v_i - v_d = \Delta v_i - \Delta v_d \, . \qquad (6)$$

The most convenient set of state variables for this problem is position error Δp, inertial velocity error Δv_i, platform tilt $\Delta\alpha$, gyro drift Δd, and scale factor error ΔSF. Written in matrix form, the differential equations for the state variables are

$$\begin{bmatrix} \Delta\dot{p} \\ \Delta\dot{v}_i \\ \Delta\dot{\alpha} \\ \Delta\dot{d} \\ \Delta\dot{SF} \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & g & 0 & 0 \\ 0 & -1/R_e & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \Delta p \\ \Delta v_i \\ \Delta\alpha \\ \Delta d \\ \Delta SF \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ 0 \\ n_d \\ n_s \end{bmatrix} . \qquad (7)$$

The measurement equation is

$$\Delta v = \begin{bmatrix} 0 & 1 & 0 & 0 & -v \end{bmatrix} \begin{bmatrix} \Delta p \\ \Delta v_i \\ \Delta\alpha \\ \Delta d \\ \Delta SF \end{bmatrix} - \epsilon_v \, . \qquad (8)$$

The filter for this probelm is now completely specified by Equation (3) where the $\vec{x}$ vector, F matrix, $\vec{n}$ vector, H matrix, and $\vec{u}$ vector are defined by Equations (7) and (8). The matrices R and Q are defined by

$$R = [\overline{\epsilon_v^2}]$$

$$Q = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & N_d & 0 \\ 0 & 0 & 0 & 0 & N_s \end{bmatrix}$$

where

$$\overline{n_d(t)\, n_d(\tau)} = N_d\, \delta(t - \tau)$$

$$\overline{n_s(t)\, n_s(\tau)} = N_s\, \delta(t - \tau)\ .$$

The initial estimates for the state variables are assumed to be zero. It is also assumed that there is no initial cross-correlation between state variables. The initial covariance matrix E is then a diagonal matrix made up of the expected mean squared value of each of the variables themselves, since the initial estimates are zero. With numbers specified for all the constants involved, all the information is now available to program the filter for this problem into a flight computer.

The results for the simulated operation of this system are given for the following typical values:

initial position error = 0

initial velocity error = 0

initial tilt = 0.2 minutes

initial gyro drift = 0.015°/hr

initial scale factor error = 0.5%

random walk gyro drift (N_d) = $(0.0015^\circ/\text{hr})^2$/hr

random scale factor (N_s) = $(0.05\%)^2$/hr

uncorrelated velocity error ($\sqrt{R}$) = 0.5 knots.

It is assumed that the vehicle accelerates to 500 knots, holds this velocity for one hour, accelerates to 1200 knots, and then stops after 4 hours. Figure 4 shows the position, velocity, and tilt error in the inertial system without the filter. Figure 5 shows the same variables with the filter. Figure 6 shows the estimates of gyro drift and scale factor error that are made by the filter. Figure 7 shows the expected uncertainty in each of the state variables. These uncertainties are given by the square root of the diagonal elements of the covariance matrix. It can be seen that there is a drop in most of the uncertainties at one hour when the velocity changed. This is because the change in velocity helps the filter distinguish between inertial errors and Doppler errors, because the Doppler errors depend on velocity while the inertial errors do not.

5. SIMULATIONS

The next set of figures shows the results of a much more realistic example. The simulation is of a supersonic transport flying on a great circle flight from New York to London. The plane flies at 500 knots over land and 1200 knots over the Atlantic. The aircraft carries an inertial system and a Doppler radar system which are combined by a statistical filter which also processes check point information. The inertial system has a variety of errors, including 0.015 deg/hour gyro drift rate and 0.1% scale factor errors in the gyros and accelerometers. The system begins navigation 15 minutes before the airplane begins moving, with 1 degree errors in vertical alignment and azimuth. Thus the navigation program also performs the initial alignment and gyrocompassing operation. The Doppler radar has a random walk scale factor error with a 0.3% initial value. When the aircraft is over water, there is a Doppler velocity error due to surface motion and due to ocean currents. The Kalman filter is not optimum, in that the scale factor errors and Doppler errors caused by operation over water are not included in the filter. These errors are absorbed by other terms. This is not the ideal situation but is realistic considering the limitations on computer size. Figure 8 shows the pure inertial error. The error which results when the inertial system is augmented by Doppler radar and filtering is shown in Figure 9. The same conditions exist in Figure 10, except that conventional filtering is used. Figure 11 shows the results when four position fixes that have a root-mean-square uncertainty of 1000 ft are incorporated into the system.

6. APPLICATIONS AND CONSEQUENCES OF FILTERING

The most important application of filtering techniques is essentially the one illustrated in the previous example, that of combining various sources of information during navigation. The Kalman filter gives a unified method of combining information from any source: inertial, Doppler, stellar, LORAN, TACAN, Decca, communications satellite, multiple inertial systems, etc. When a statistical filter is used with an external velocity source, two of the classical problems in inertial navigation - damping and gyrocompassing - are automatically solved in an optimum way. The 84-minute and 24-hour period oscillations of the inertial system errors are described by the error equations which are written into the filter. In fact, these characteristics are one of the primary means by which the filter distinguishes inertial errors from other errors. Thus, these error oscillations are automatically "damped". Also, the filter uses all available information to estimate the position and velocity of the vehicle. This includes obtaining an indication of the direction of North by measuring the rotation of the gravity vector that is due to the rotation of the earth. Thus, a system with a filter is continuously "gyrocompassing". Kalman filtering can be used for optimum damping and gyrocompassing even with no external reference. (See Reference 4). However, even optimum damping is probably not very effective for an aircraft inertial navigation system.

Another classical problem which is greatly aided by statistical filtering techniques is the initial alignment and calibration of an inertial system. A filter of the same form as that used for navigation can be used, during pre-flight, to process external information in order to estimate the misalignments and miscalibrations of the system. The external information may be just the fact that the aircraft is not moving, so that any bias velocity is velocity error. The state variables that are used in the pre-flight filter normally include platform misalignments, gyro drifts, and any other terms which might be subject to day to day variations, such as gyro and accelerometer scale factors. In addition, the calibration program might be run with the platform in several different attitudes to help the filter separate error sources.

The calibration problem and the navigation problem are so similar that in many cases the same computer program can be used. It is necessary to include in the navigation filter most of the major inertial system error sources, such as misalignments and gyro drifts. These parameters are the same ones that must be estimated during the pre-flight calibration program. Thus, with only quantitative changes in some of the statistical values describing the system, the navigation filter will also handle the calibration problem. For example, a filter which is designed to use Doppler radar measurements can be used for calibration by making the Doppler measurement zero and by greatly reducing the assumed error in the measurement.

In most cases, filtering techniques will both improve the accuracy of the calibration process and shorten the time required. A filter can easily handle the problem of motion during the alignment, such as wind buffeting, and can also be mechanized to account for the changing characteristics of components during warm-up. The same type of filter can be used to transfer alignment from a master system to another inertial system, such as in aircraft carrier operations.

Filtering techniques are also useful for post-flight analysis of navigation systems during a test program. In this application, all necessary system data is recorded during flight, along with all available reference data. The reference data is used in the same way as it is used in flight; but reference data is usually much more accurate and the error model for the navigation system can be much more complete. In post-flight analysis even better use of the information can be made by extending the filter to include optimum smoothing techniques, which means that reference data is used from both sides of the point in question.

One of the most important consequences of the use of statistical filtering for navigation is that it changes the basic figure of merit for the component parts of the navigation system. The dynamic and statistical nature of the error becomes very important in addition to its absolute accuracy. What is ultimately important is the total accuracy of the overall system after the measurements have been processed by the filter. Thus, what is important is the ability of the filter to detect and compensate for the errors. For example, if a component has poor day-to-day stability but the error is nearly constant once warmed up, the total system accuracy might be better than with a component with better absolute accuracy but with less stability during the flight. The parameters that give a measure of this lack of stability during flight are the white noise inputs which drive the error equations. These are given by the Q and R matrices in the filter equations (see Equation (3)). Thus, one of the important parameters of a component is the size of this white noise term that is a part of the error model for that component. For example, if gyro drift can truly be described by a random walk model as used in the example in Section 4, then an important figure of merit for a gyro is the increase in mean squared gyro drift per unit time. The units for this parameter would be $(\text{deg/hour})^2/\text{hour}$. When a filter is used, this number might be more important than the drift rate itself.

An even more important consideration is the dynamic characteristics of the errors. If a filter is used to combine the information from two systems which have errors with the same dynamic characteristics, then the filter can do little more than produce a statistical average which might reduce the error around 20 to 30%. On the other hand, if the errors have unique dynamic characteristics, then the filter will be able to distinguish the errors and obtain a possible 70 to 80% improvement. As was mentioned in the examples, this situation is an advantage when an inertial system and Doppler radar are combined. Most inertial errors are related to some inertial direction while Doppler errors are related to aircraft axes. Thus, when a turn is made, the filter is to some degree able to distinguish and calibrate the errors.

A very valuable by-product of the use of statistical filtering is that an *a priori* estimate is available at all times for the accuracy of the navigation system. This feature has at least three possible uses. First, the information could be displayed directly. The navigation system would then give the vehicle operator not only the indicated position and velocity, but the estimated accuracy of these numbers. In some situations this information could be invaluable. Secondly, a record could be kept of the actual and estimated error in the system, for example, at terminal points. If there were, on an average, too much discrepancy between how well the system was doing and how it thought it was doing, this would indicate that the system was not operating properly. The problem could be due either to some component which was exceeding specifications or to improper modeling of the system. A third use of this statistical data would be the automatic editing of the input data. The filter already has an estimate of the expected error in the measurements. Thus, if a measurement is in error by more than three or four standard deviations, it can be automatically rejected and indication can be given to the operator that something might be wrong.

7. PROBLEMS

Two of the most important problems in the use of statistical filtering are the requirements for a large flight computer and the need for adequate statistical models of the component parts of the system.

The computational problem is due to the necessity of computing, in real time, the optimum statistical weighing factors. The weighing factors involve the integration of a matrix differential equation for the covariance matrix (see Equation (3)). This matrix has dimensions $n \times n$, where n is the number of variables being estimated by the filter. The amount of computation required is roughly proportional to n^3. Thus, the more complete the model, the worse the computational problem. For a particular navigation system there will be a trade-off between accuracy and computer capability something like the hypothetical curve shown in Figure 12. For any system there will be a point beyond which a larger computer gives very little improvement. Considerable knowledge of the system is needed to actually determine the trade-off between accuracy and computer size. This fact leads into the other major problem.

The other major problem is the need for statistical models for all the instruments involved in the navigation system. Computer simulations of typical navigation problems using supposedly realistic statistical models produce outstanding performance when Kalman filtering is used. But the results in the real world are not necessarily this good, because of unsuspected errors with which the filter is not capable of coping. A detailed statistical error model for an inertial system has not really been essential in the past, but now that the model is actually a part of the system, its determination is much more important. To be confident of having a complete model, it is necessary to test the system in an operational environment as near as possible to the one in which the system is actually going to be used. This flight test program is necessary to assure that the statistical filter can handle any peculiarity in the system. It is very helpful in this flight test program to record all information that is necessary to "re-fly" the flight on a ground-based computer. With this recording it is possible to change the filter and determine what the results would have been without having to re-fly the aircraft. It is thus possible to optimize the filter with great savings in flight test expenses.

8. CONCLUSIONS

It is very likely that a large percentage of the aircraft in the future will use inertial navigation equipment and that the mechanization of these systems will be greatly influenced by statistical filtering techniques. These techniques are in an early stage of development and much experience is needed to determine just how they should be used. In many cases, it is likely that a simplified filter will produce results which are nearly optimum. On the other hand, new developments in very highly capable, but less expensive, flight computers may make desirable even more sophisticated filtering techniques. In either case, the Kalman filter equations give a unified foundation from which practical mechanizations can be developed.

REFERENCES

1. Kalman, R.E.	*A New Approach to Linear Filtering and Prediction Problems.* Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers, March 1960.
2. Kalman, R.E. Bucy, R.S.	*New Results in Linear Filtering and Prediction Theory.* Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers, March 1961.
3. Wiener, N.	*The Extrapolation, Interpolation and Smoothing of Stationary Time Series.* John Wiley, New York, 1949.
4. Brock, L.D.	*Application of Statistical Estimation to Navigation Systems.* Ph.D. Thesis, T-414, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, June 1965.

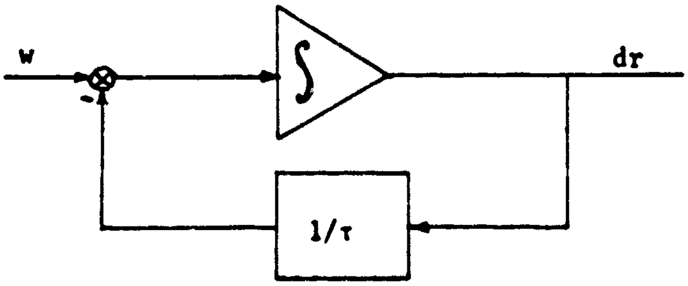


Fig.1 Shaping filter for stationary gyro drift rate

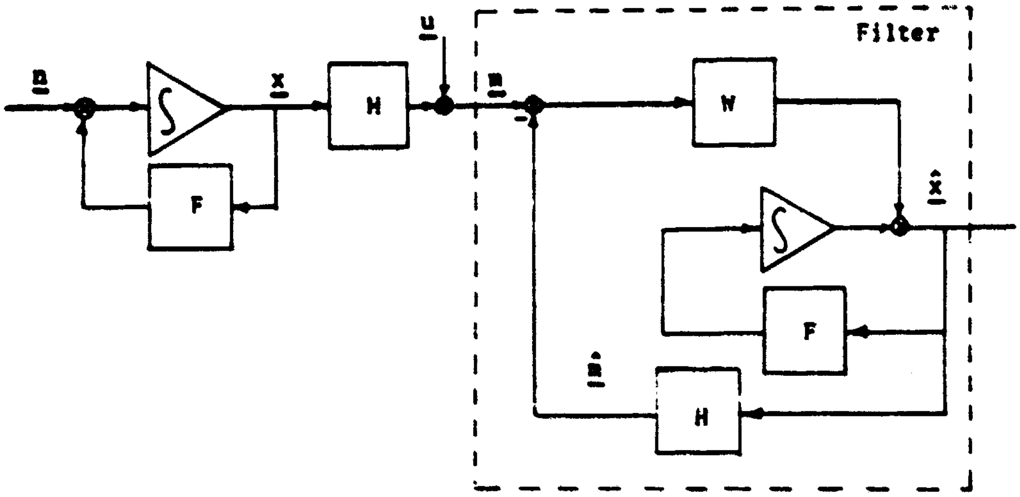


Fig.2 System with Kalman filter

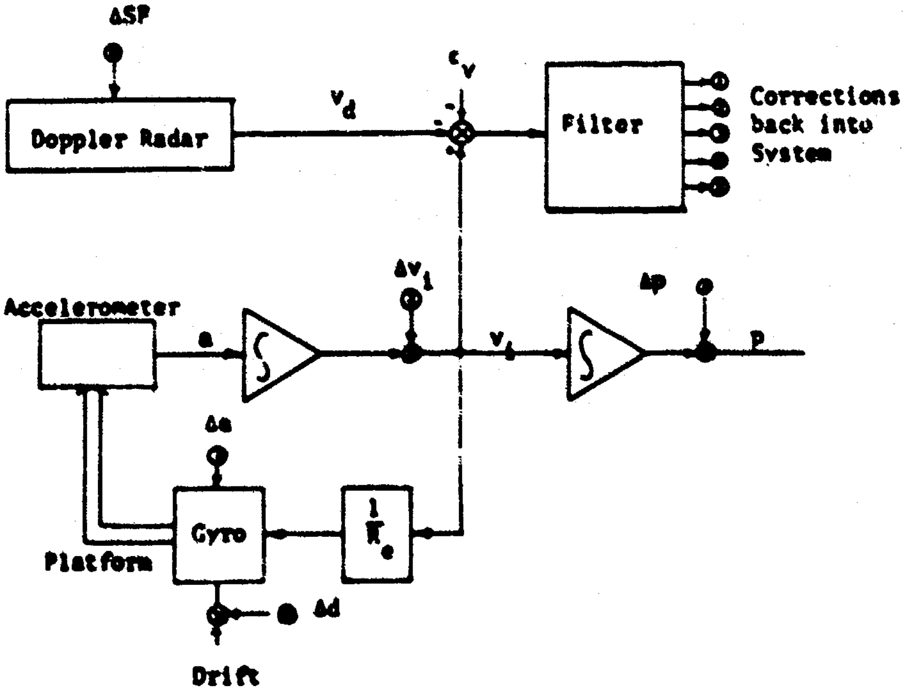


Fig.3 Example system

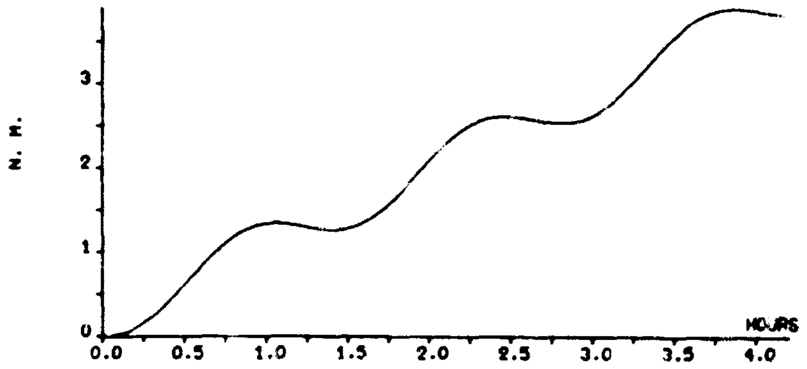


Fig.4(a) Pure inertial position error

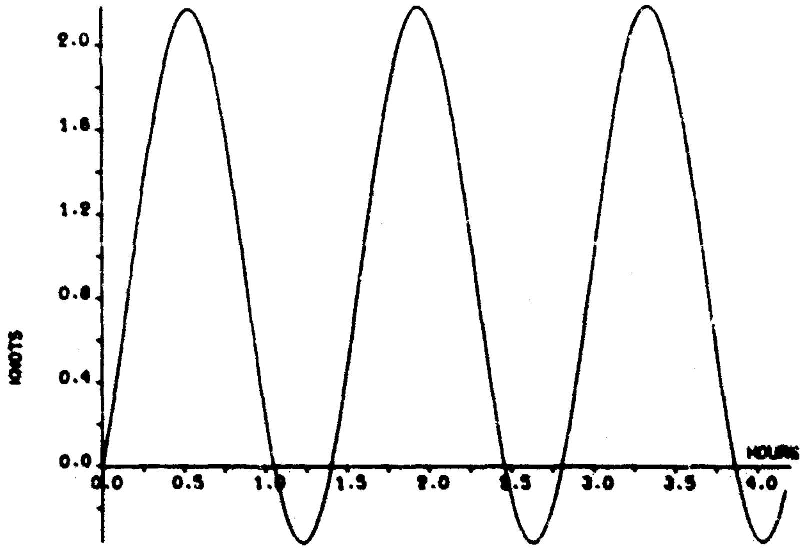


Fig.4(b) Pure inertial velocity error

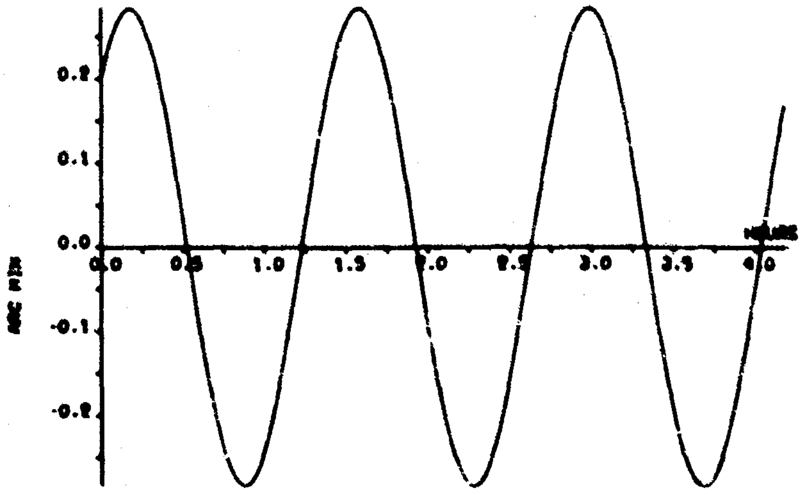


Fig.4(c) Pure inertial vertical error

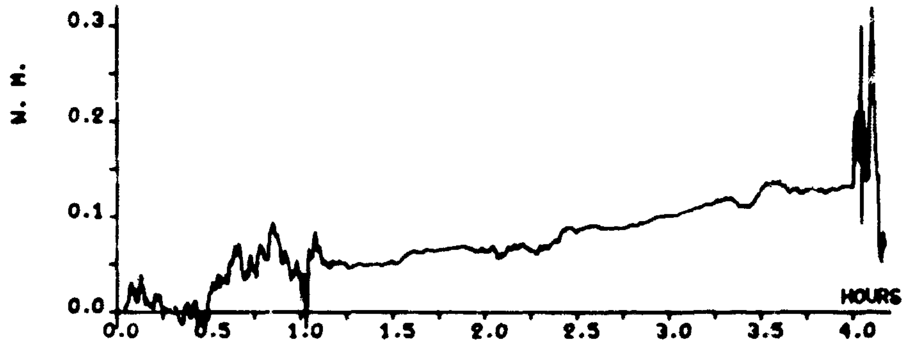


Fig. 5(a) Position error with filter

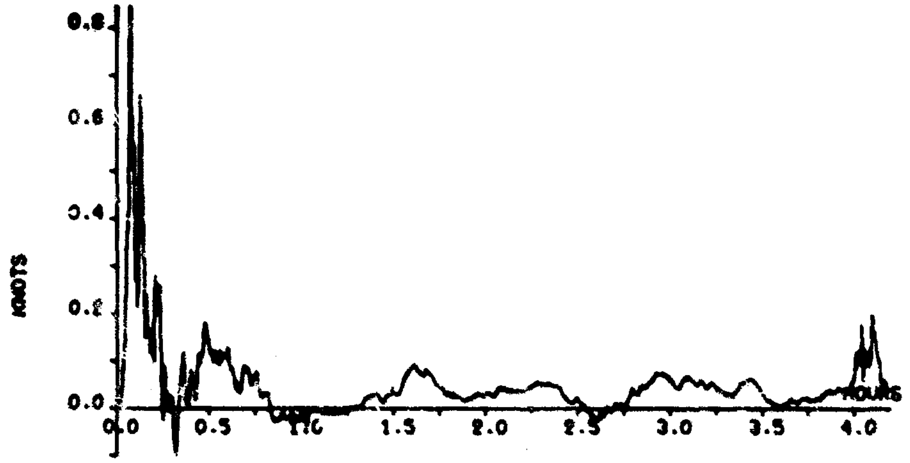


Fig. 5(b) Velocity error with filter

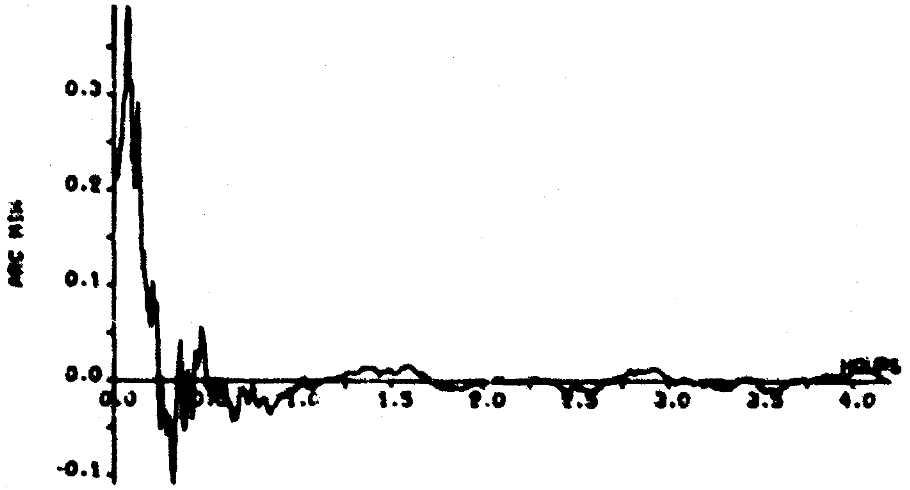


Fig. 5(c) Vertical error with filter

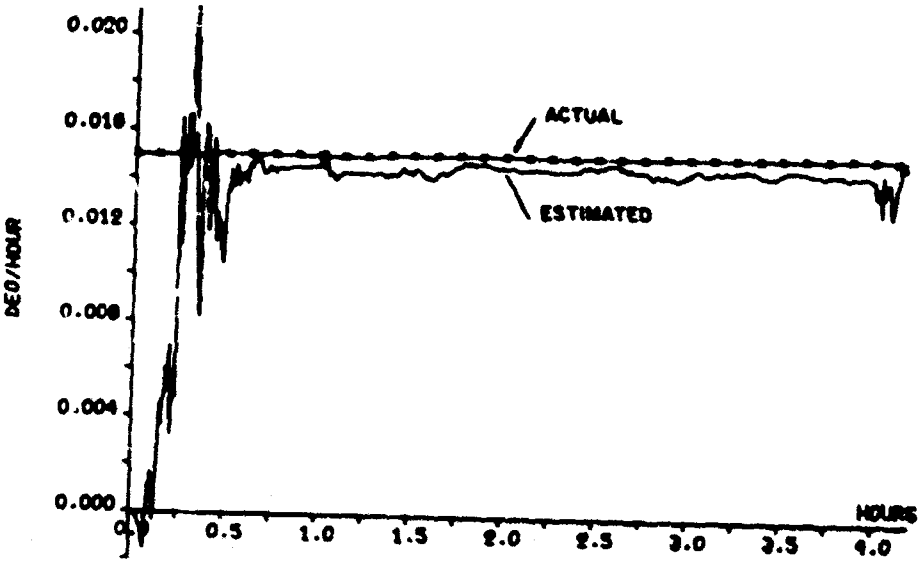


Fig.6(a) Actual and estimated gyro drift

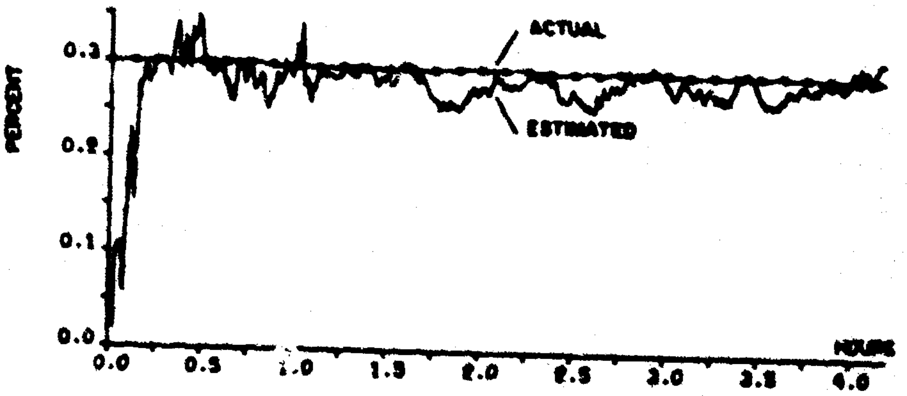


Fig.6(b) Actual and estimated scale factor

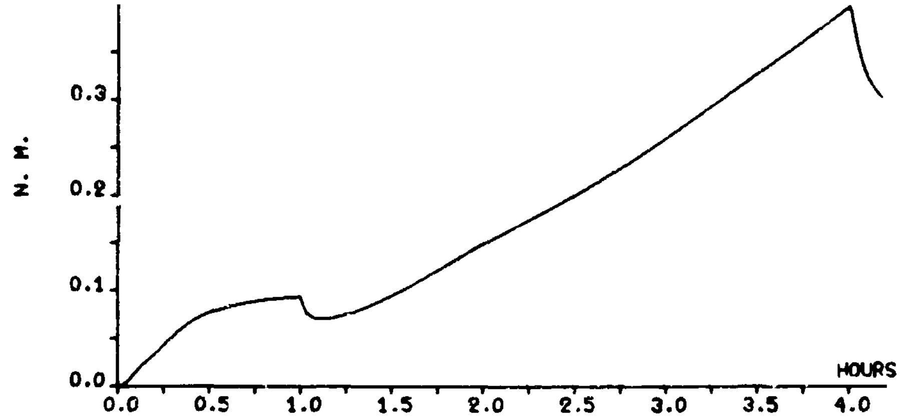


Fig.7(a) Root mean square position uncertainty

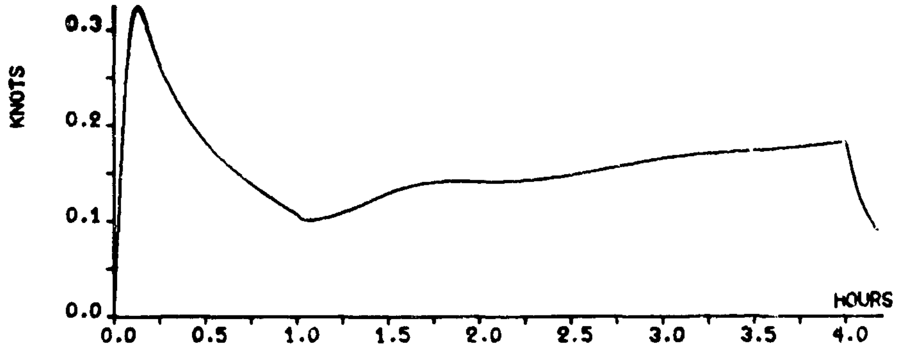


Fig.7(b) Root mean square velocity uncertainty

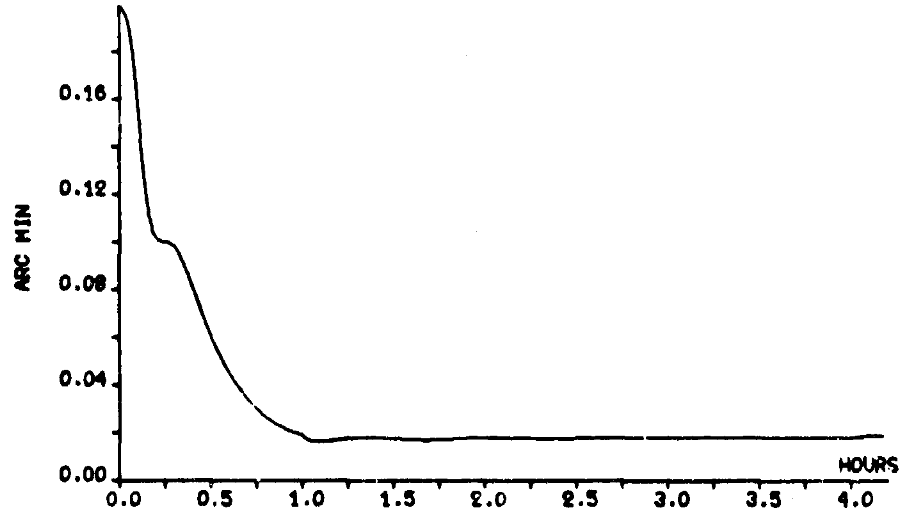


Fig.7(c) Root mean square vertical uncertainty

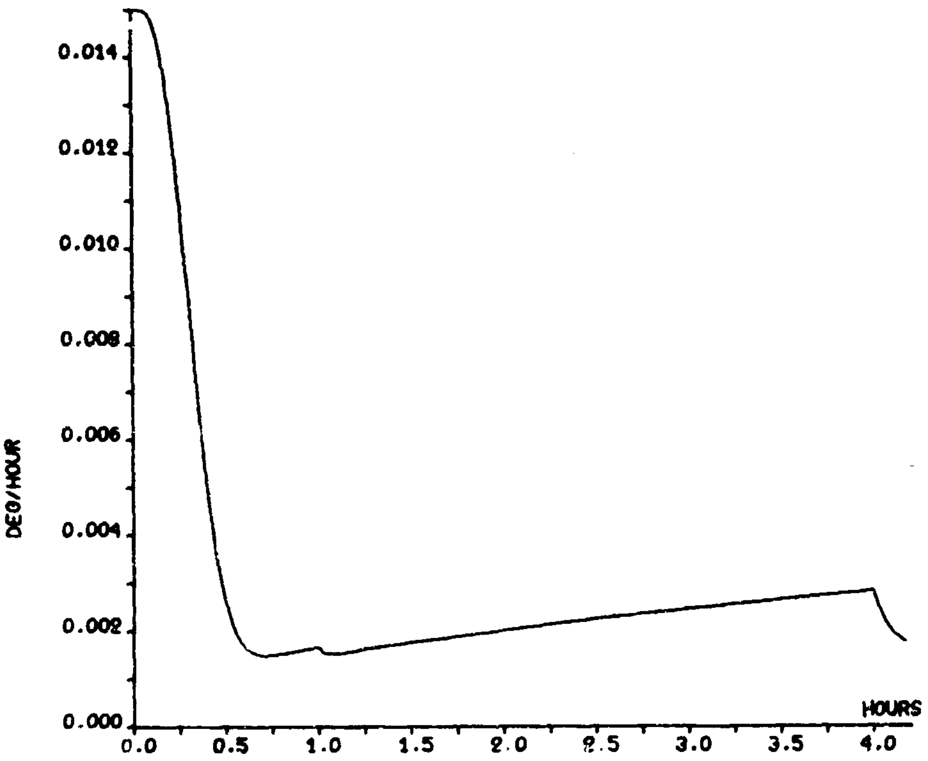


Fig. 7(d) Root mean square gyro drift uncertainty

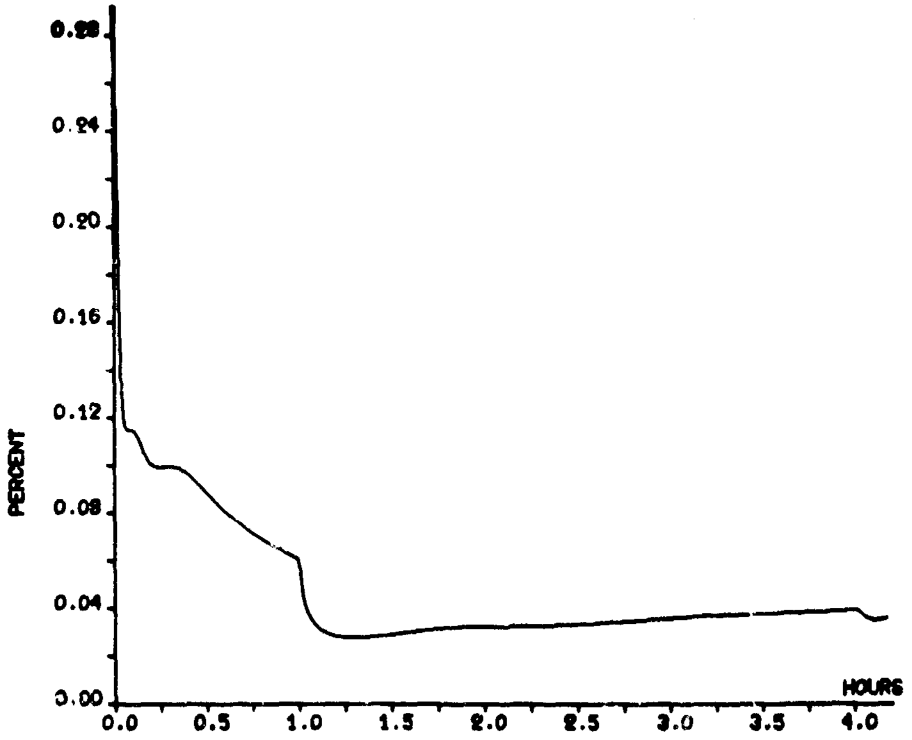


Fig. 7(e) Root mean square scale factor uncertainty

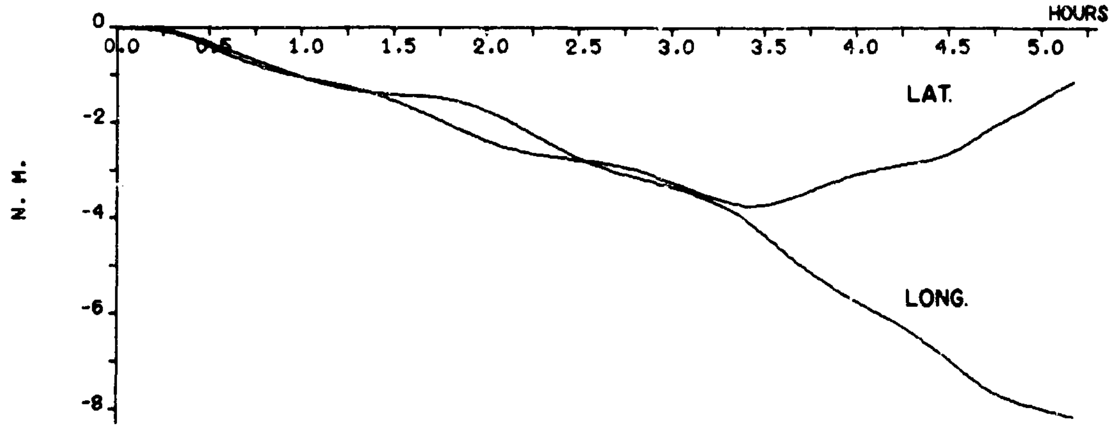


Fig. 8 Pure inertial position error

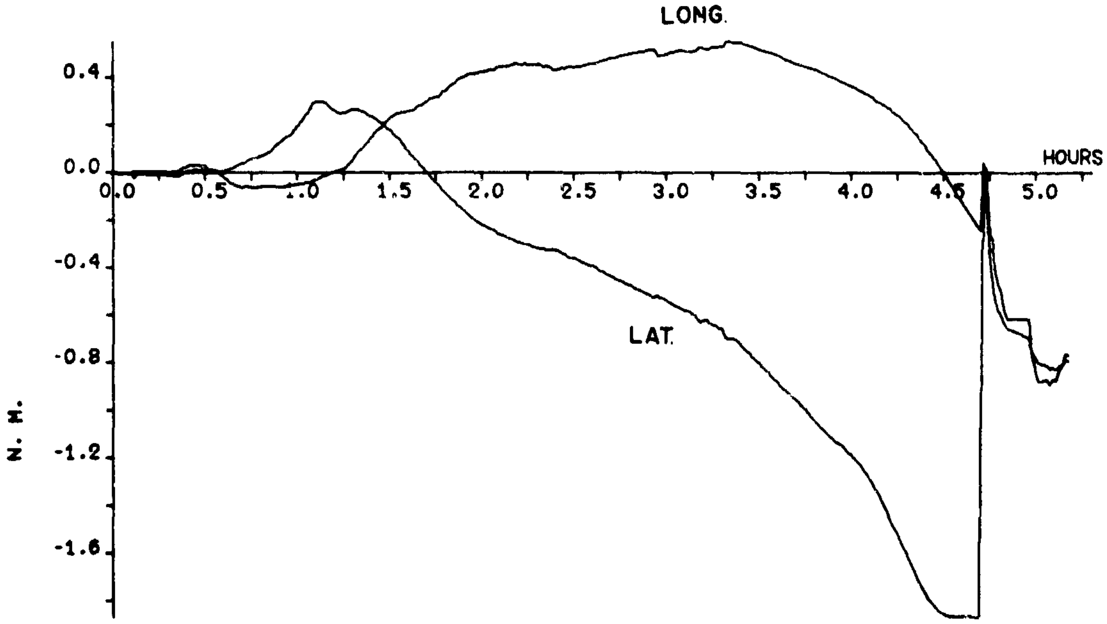


Fig. 9 Position error with filter

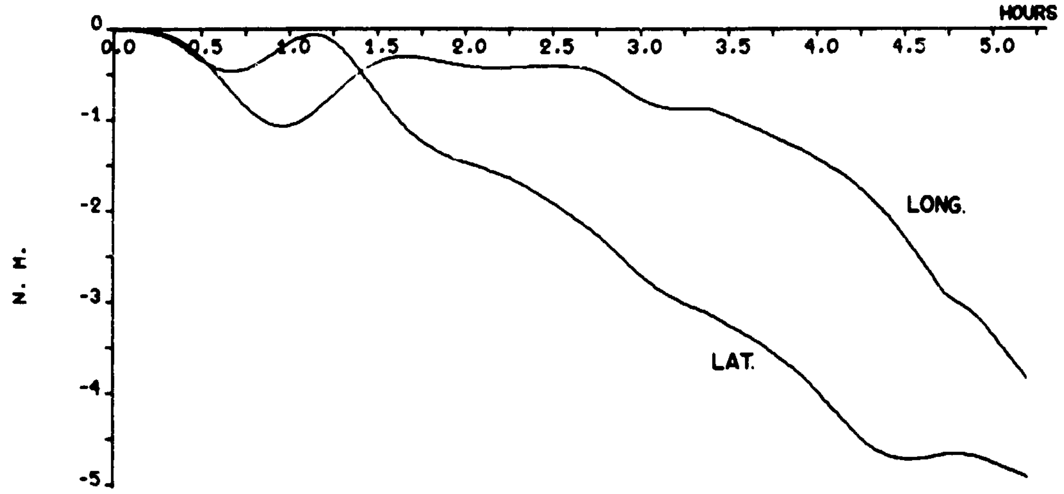


Fig.10 Error with conventional filtering

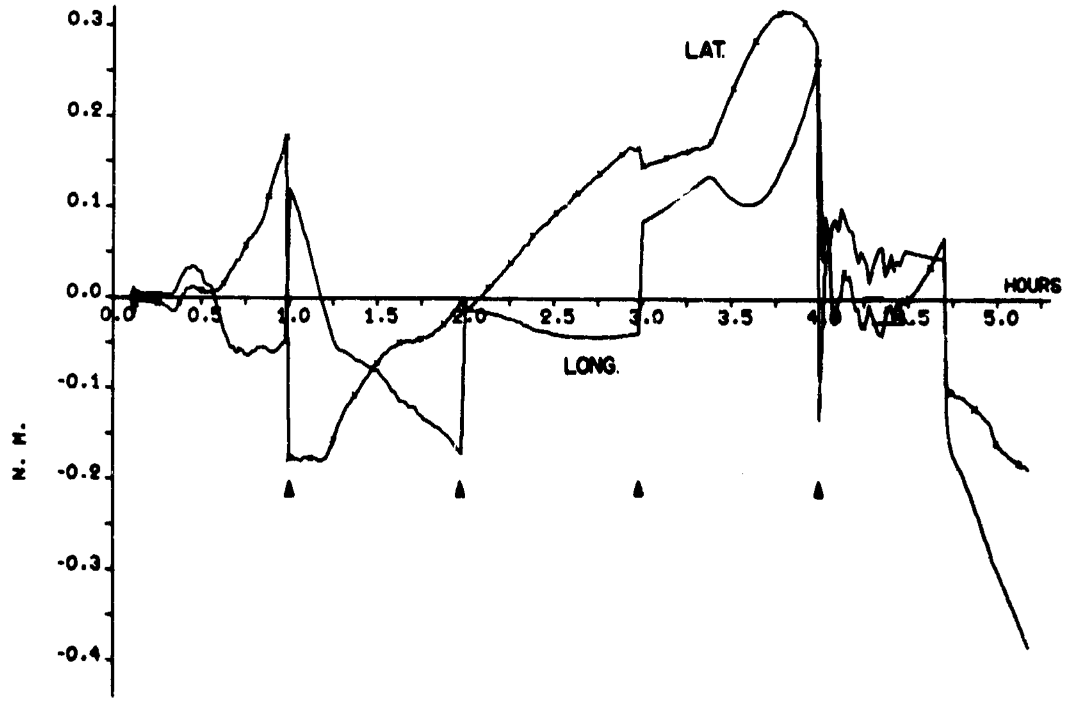


Fig.11 Error with position fixes

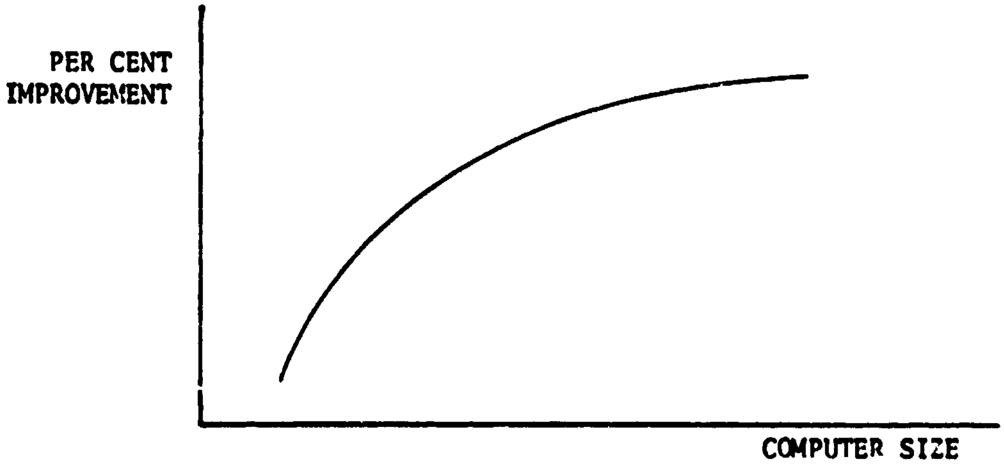


Fig. 12   Trade-off between accuracy and computer size

APPLICATION OF INERTIAL TECHNOLOGY TO AIRBORNE GRAVIMETRY

by

Elmer J.Frey and Raymond B.Harlan

Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

SUMMARY

Under sponsorship of the United States Air Force Cambridge Research Laboratories the Massachusetts Institute of Technology undertook to apply the techniques of inertial guidance to airborne gravimetry. Tests of a missile accelerometer used as a gravimeter on a stabilized platform aboard a KC135 aircraft ensued.

The results of an error analysis of the stabilization and navigation requirements are presented. Navigation requirements are divided into horizontal and vertical coordinates, in the horizontal coordinates accurate velocity is shown to be the most demanding requirement. In the vertical coordinate, filtering is required to distinguish between gravitational and inertial reaction forces.

The design of the experimental gravimetric system is then discussed, with emphasis on the pendulous gyro gravimeter and on the stabilization system.

Some results of the flight tests, which started in November 1966, are presented in the form of raw data and reduced observations. Incidental data involving air turbulence and aircraft motion are also presented.

NOTATION

a	semi-major axis of reference ellipsoid	cm
a_z	acceleration along the local vertical	cm/sec^2
E	Eötvös correction	cm/sec^2
f	earth's flattening (1/298.3)	
f_N, f_E	specific force along North, East	cm/sec^2
f_y	specific force along y aircraft coordinate	cm/sec^2
h	altitude of aircraft above reference ellipsoid	cm
V	aircraft groundspeed	cm/sec
γ_0	gravity at sea level	cm/sec^2
λ	longitude	rad
ϕ	geographic latitude	rad
ω	earth rotation rate (7.29×10^{-5})	rad/sec
ψ	ground track angle measured from north	rad
Δ	used as prefix to denote increment or error in a quantity	

APPLICATION OF INERTIAL TECHNOLOGY TO AIRBORNE GRAVIMETRY

Elmer J. Frey and Raymond B. Harlan

1. INTRODUCTION

Gravimetry is the measurement of the force of gravity and was first done in static conditions on the surface of the earth. The instruments used for this purpose have principally involved measurement either of the period of a pendulum or of the distortion of an elastic spring[1]. During the 1920s Vening Mainesz[2] used pendulum apparatus to measure gravity at sea aboard a submerged submarine, and in the following decade spring-type gravimeters were designed for use in relatively static conditions at the sea bottom at modest depths[1]. In 1957 Worzel[3] measured gravity at sea on board a surface ship using the spring-type meter of Graf, and not too much later the elastic spring meters of LaCoste-Romberg were also successfully used on surface ships at sea. Finally, in 1959, the first experiments in airborne gravimetry were conducted by Thompson and LaCoste[4]. These were soon followed by other experiments in airborne gravimetry, notably those of Nettleton, LaCoste, Glicken, and Harrison[5,6], always using elastic spring gravimeters. Thus, the measurements have progressed from static conditions to increasing levels of velocities and accelerations.

A gravimeter measures specific force, and is thus identical in function to an accelerometer, except for the anticipated environment. A seismometer may also be considered to have an identical function, and in fact several of the current gravimeters, such as the LaCoste-Romberg, are adaptations of the LaCoste seismograph suspension system. The standard unit of the geodesist is the milligal, with a value of 0.001cm sec^2 or approximately 10^{-6} gravity; a land gravimeter is expected to have an accuracy of at least 0.1 milligal, or approximately 10^{-7} gravity, and instruments with considerably higher resolution have been used for many years. However, such units must be clamped during transportation, and usually have a very limited range of static measurement typically a few thousand milligals, and sometimes less than 50 milligals. Exposure of an uncaged instrument to conditions beyond this range necessitates recalibration and may produce permanent damage. Therefore, the instruments designed for a dynamic environment have sacrificed some of the resolution and accuracy of the static instruments in order to survive and measure in the dynamic environment. Nevertheless, the surface ship and airplane measurements mentioned thus far all employed instruments which were basically adaptations of static land gravimeter designs.

The measurement of gravity on a moving vehicle introduces other problems besides those of the gravimeter; for example, it becomes necessary to provide angular stabilization or indication in order to point the sensitive axis of the gravimeter in the proper direction, i.e., along the gravity vector. In addition, it becomes necessary to distinguish between the accelerations of gravity and vertical accelerations with respect to the earth, both of which affect the instrument equally, by some appropriate filtering process. Finally, it is also necessary to distinguish the vertical Coriolis accelerations due to horizontal velocities, which implies that these velocities must also be known. These requirements further limit the accuracy and the resolution of gravimetry on board a moving vehicle, particularly on an aircraft where altimeter accuracy limits the knowledge of the vertical accelerations as well as of the altitude itself. Consequently, even a perfect gravimeter with infinite resolution can provide only results consistent with the other sources of inaccuracy.

Since the measurement of specific force, angular stabilization, and indication of position and velocity are all the subject matter of inertial navigation, the application of the techniques and equipment of this field to airborne gravimetry is logical. In 1965 the Experimental Astronomy Laboratory started a study of the possibility of using inertial equipment for airborne gravimetry. The study led to the selection of an accelerometer for test as a potential gravimeter, and the assembly of a stabilization platform for the gravimeter for a flight test program. By late 1966 experimental gravimetric flights were being conducted in a United States Air Force KC-135 aircraft, using this equipment and the navigation system which already existed in the aircraft.

2. STABILIZATION AND NAVIGATION SYSTEM ANALYSIS

2.1 Stabilization

A specific force along any direction can be computed from force measurements made in three arbitrary, but known, orthogonal directions. Thus in principle a "strapdown" measurement system of three instruments could be used for gravity measurement. Such an arrangement requires three accurate instruments rather than one, and in addition requires very high accuracy in the determination of the direction of the input axes of the instruments and stability in these directions with respect to each other. A typical error coefficient for alignment uncertainty is three milligals per arc second. Consequently, it is preferable to use one instrument aligned along the desired direction, because the courses flown can be rather smooth, and accelerations normal to the desired direction may be kept small. This reduces alignment errors to a second-order effect.

If a pendulous gyro accelerometer is used for a gravimeter, an additional requirement is made; since the output of the instrument is a rotation about the input axis, it must either be stabilized about the vertical axis or the azimuth motion of the case must be known.

Examine first the alignment error for a static gravity measurement on land or at the sea bottom. In this case the instrument is to be aligned along the gravity vector, and there is no horizontal acceleration. Thus alignment may be done by levels, and an error $\Delta\theta$ in verticality produces an error $\frac{1}{2}g(\Delta\theta)^2$ in the measurement, so that a one milliradian error in leveling produces about one half milligal error, and a one arc minute error produces about 0.04 milligal. Land and sea bottom gravimeters are usually leveled by static methods.

Early submarine and shipboard gravimeters were leveled by use of pendulous gimbal systems, with periods of the order of one or two minutes. Since the pendulous gimbals followed the apparent vertical, i.e. the specific force vector, the resulting measurements had to be corrected for the horizontal accelerations which were sensed and for the second-order reduction in the gravity vector sensed. This latter computation, called the Browne correction[7], is about 50 milligals for a horizontal acceleration of 0.01g and increases with the square of the acceleration. The analog computation commonly used to compute this correction was also used to make partial correction for the dynamic response of the gimbals[3,8,9]. Subsequently gyro-stabilized platforms were used to replace the pendulous gimbal systems, reducing considerably the effect of horizontal accelerations[10-14].

In analyzing the stabilization requirement, the first step is to express the horizontal specific force of the vehicle in a suitable coordinate system. In horizontal flight, neglecting all terms in the earth's flattening reduces the expressions for North and East specific force to[15]

$$f_N = -a\ddot{\Phi} - a\dot{\lambda}(\dot{\lambda} + 2\omega)\sin\Phi\cos\Phi \qquad (1)$$

$$f_E = -a\ddot{\lambda}\cos\Phi + 2a\dot{\Phi}(\dot{\lambda} + \omega)\sin\Phi\ . \qquad (2)$$

Note that the centripetal term $a\omega^2 \sin \Phi \cos \Phi$ is part of the gravity vector and thus does not appear in the horizontal acceleration expression.

Since gravimetric survey flights are likely to follow simple patterns, it is convenient to look at components of horizontal specific force for great-circle and rhumb-line flights at constant speed. Along the flight direction, the specific force is zero; normal to the flight direction, for great-circle flights, it is

$$f_y = 2V\omega \sin \Phi , \tag{3}$$

and for rhumb-line flights it is

$$f_y = 2V\omega \sin \Phi + \frac{V^2}{a} \tan \Phi \sin \psi . \tag{4}$$

For great-circle courses at 400 knots, the Coriolis acceleration of $2V\omega \sin \Phi$ can be as much as 3000 milligals, and is the dominant sustained horizontal acceleration. At 400 knots and 90° or 270° heading, the additional rhumb-line acceleration reaches the same value at latitudes above 75°. Thus a reasonable estimate for sustained mean horizontal acceleration for all flights is 3000 milligals; this is the cross-acceleration term which produces alignment errors if the gravimeter is to be oriented along the gravity vector. The error sensitivity in this case is approximately one milligal per arc minute of tilt in the roll direction.

Several possibilities for stabilization coordinates appear. One is the usual North, East, and vertical system commonly used in navigation platforms. Another is a smoothed or averaged apparent vertical such as is provided by an aircraft vertical gyro reference; this has the disadvantage of being sensitive to horizontal accelerations of short duration, which means that both a Browne correction and a cross-acceleration correction must be applied. The cross-acceleration correction may be computed from navigation information and knowledge of the dynamic response of the stable element, but involves inaccuracies arising from these sources[9,11].

Another coordinate system which follows the apparent vertical for flight paths unaccelerated with respect to the rotating earth is possible, and would present the advantages of making the cross-accelerations for such flights approximately zero. Such a coordinate system is that which a local vertical inertial navigation platform would follow if the Coriolis acceleration compensations were omitted. The Browne correction would again be required, but the platform attitude could be more easily determined and the Browne correction could be computed rather easily and accurately. The error sensitivity for such a configuration would depend primarily on rhumb-line accelerations, and a value of 0.3 milligal per arc minute might be appropriate as an average error coefficient.

Stabilization instrument requirements depend in part upon the course flown. Table I shows the peak error sensitivities obtained in a simulation of a three axis local vertical platform flown over the courses shown in Figure 1. The table shows angular errors in tilt of the vertical about North and East, and angular velocity errors about the azimuth axis. The latter are shown because the output of a gyro accelerometer is an angular velocity, about the input axis, proportional to the specific force input. Hence the azimuth angular velocity errors are critical when such an instrument is used as a gravity meter. Figure 2 illustrates the behavior of the angular errors due to a constant gyro drift rate, for an easterly flight along the 75° parallel of latitude. The high latitude emphasizes azimuth drift; the 84-minute period oscillation is also clearly evident for all three angular errors shown.

Figures 3 and 4 show azimuth error due to a constant drift rate in the azimuth gyro, for great-circle courses passing within 15° of the pole. The effect of high latitude is clearly evident; the simulation also showed that the time of passage nearest the poles is also important. These error coefficients are important to the case when the pendulous gyro accelerometer is used as a gravimeter; with a typical accelerometer scale factor of about one radian per second per gravity, an azimuth gyro drift rate of one minute of arc per hour can produce a gravimeter output error of 0.8 milligal for parts of the course shown in Figure 3.

The error sensitivities for platform levelling which are shown in Table I are not especially demanding as far as the current state of the art is concerned, if one adopts a one minute of arc specification, which corresponds to one milligal gravimeter error. However, it should be noted that reasonable velocity signals are necessary to keep the levelling this accurate, since four knots of velocity produces almost one arc minute of tilt.

The preceding analysis covered the case of the platform normally aligned to local North, East, and vertical. Time did not permit carrying out a similar, complete analysis for the case of a platform aligned to the difference between the gravity vector and the aircraft Coriolis acceleration vector, i.e. to the apparent vertical for flight with no horizontal accelerations with respect to the rotating earth. However, since the two coordinate systems are fairly close to coincidence, within 20 arc minutes of each other, and since the platform dynamics do not differ markedly, the error coefficients should be approximately the same.

2.2 Navigation Requirements

The navigational requirements may conveniently be separated into those relating to the horizontal coordinates, or latitude and longitude, and those relating to altitude, since the way in which the two different coordinates affect airborne gravimetry is drastically different. Since the flight test program was based on use of an existing navigation system, the analysis of navigation errors was more limited in scope than that for stabilization errors.

2.2.1 Navigation-Horizontal Coordinates

Mapping the gravity anomalies requires knowledge of the location of each observation, i.e. navigation information. The gravity anomaly is the difference between the actual and the nominal gravity value at each point, and the nominal value is a function of both latitude and altitude. The standard sea level gravity value as a function of latitude is given by Heiskanen as

$$\gamma_0 = 978.049(1 + 0.0052884 \sin^2\Phi - 0.0000059 \sin^2 2\Phi) . \tag{5}$$

The change with latitude is not significantly different at aircraft altitudes and for purposes of error analysis the sea level formula is adequate. The rate of change of standard gravity with latitude is

$$\frac{d\gamma_0}{d\Phi} = 978.049(0.0052884 \sin 2\Phi - 0.0000118 \sin 4\Phi) , \tag{6}$$

which takes on its maximum value near 45^0 latitude, where the change is approximately 1.5 milligal per arc minute. This ratio provides one error criterion for latitude indication. The error criterion for longitude indication depends on the resolution desired for the map. The resolution attainable for airborne gravimetry is limited for other reasons which are discussed in more detail below; these limitations are such that the horizontal position errors will not be discussed further at this point.

Horizontal velocity is required as well as position, since the gravimeter is sensitive to the vertical Coriolis and centripetal accelerations. A sufficiently accurate expression for the component of acceleration of a body along the normal to the earth ellipsoid is[16]

$$a_z = a\dot{\Phi}^2\left[1 + \frac{h}{a} + f(3\sin^2\Phi - 2)\right] + a(\dot{\lambda} + \omega)^2\cos^2\Phi\left[1 + \frac{h}{a} + f\sin^2\Phi\right] - \ddot{h} . \tag{7}$$

One part of this expression is the centripetal acceleration of a body fixed with respect to the rotating earth; this is the term

$$a\omega^2\cos^2\Phi\left[1 + \frac{h}{a} + f\sin^2\Phi\right] , \tag{8}$$

which is included in normal gravity and has just been discussed. The remaining terms involve vehicle motion with respect to the earth, and must be removed from the gravimeter reading to leave the gravity observation. Since this portion of the discussion is devoted to horizontal motion, the term $\ddot{h}$ will be ignored and it will be assumed the altitude is constant. The remaining terms are known to the geophysicists as the "Eötvös correction".

For the purposes of an error analysis, the small terms with coefficients of h/a and f which appear within the brackets in Equation (7) may be ignored. The resulting simplified expression takes the form

$$E = a\dot{\Phi}^2 + a(\dot{\lambda}^2 + 2\dot{\lambda}\omega)\cos^2\Phi . \tag{9}$$

The corresponding error expression is

$$\Delta E = 2a\dot{\Phi}\Delta\dot{\Phi} + 2a(\dot{\lambda} + \omega)\cos^2\Phi\Delta\dot{\lambda} - a\dot{\lambda}(\dot{\lambda} + 2\omega)\sin 2\Phi\Delta\Phi . \tag{10}$$

The effect of latitude error $\Delta\Phi$ may be evaluated by comparison with the size of the centripetal term appearing in normal gravity; it obviously varies with longitude rate. At 45^0 latitude and 400 knot speed, the error is approximately 2.6 milligals per arc minute of latitude for eastbound flight, and 1.3 milligals per arc minute for westbound flight.

The term due to error in latitude rate contributes an error proportional to the latitude rate; at 400 knots a one knot error in velocity contributes about 3.2 milligals. The term due to error in longitude rate depends both on latitude and on longitude rate; the error coefficient increases for eastbound flight and decreases for westbound flight. At 45^0 latitude and 400 knot speed, the approximate error coefficients are 8.5 milligals per knot of error eastbound and 1.8 milligals per knot westbound.

Since the experimental program was based on the use of an existing navigation system coupled with the use of aerial photography, there was no need for a detailed error analysis for design purposes. The error coefficients just developed provide an adequate description for estimation purposes in data reduction.

2.2.2 Navigation-Vertical Coordinates

A first requirement for altitude indication arises from the fact that the gravity field is a function of altitude. Near the surface of the earth, the field attenuates by one milligal in approximately three meters. This provides the error sensitivity for the static effect.

In addition, there is a dynamic effect due to the response of the gravity meter to specific force, which includes vertical accelerations as well as gravitational accelerations. If the altitude is known with adequate precision vertical acceleration can be obtained from it by differentiating twice, and the instrument readings corrected for the acceleration term ($\ddot{h}$ in Eq. (7)). However, since differentiation of observed data introduces noise into the computed acceleration, care must be exercised in the use of the altitude data, or the data must be extremely precise.

In gravimetric flights vertical motion constitutes a disturbance whose spectrum must be examined if it is to be measured and its effects eliminated. The sources of the vertical accelerations include, among other things, atmospheric turbulence, aircraft dynamics such as phugoid motion, and variations in aircraft weight, engine performance, trim, and control surface motion. The order of the problem can be illustrated by an example of phugoid motion, which for a KC-135 aircraft has a period of about 90 seconds. A phugoid of one centimeter in amplitude creates an acceleration of five milligals in amplitude for this case. Higher frequency motions with the same amplitude in displacement create even larger accelerations.

In typical quiet air conditions, the one sigma value of observed accelerations is generally of the order of two or three thousand milligals. In turbulent air, this figure rises to values of ten or twenty thousand milligals or even more. Using the altitude hold system of the KC-135 aircraft, the vertical accelerations have a spectrum which, in conditions of turbulence, peaks in the range from 0.03 to 0.06 c/s. In quiet air the spectrum is much broader, with a relatively flat shape from about 0.01 to 0.12 c/s.

It is instructive to examine what altitude variations correspond to the observed accelerations. Choosing 0.05 c/s as the center of the spectrum, and 5000 milligals as a peak amplitude for a sinusoid, yields a peak amplitude of motion of 50 centimeters. Thus to be able to remove such an oscillation with an accuracy of ten per cent requires measuring altitude changes to an accuracy of five centimeters. Since no existing altimeter provides altitude with accuracy anywhere near this level, the readings must be averaged over a long enough time to yield suitable results. This limits the spatial resolution of airborne gravimetry since during an averaging period the aircraft moves a considerable distance.

The preceding example emphasizes the importance both of altimetry and of the filtering methods used to eliminate the vertical accelerations from the observed data. If, for example, the observed altitude variation of 50 centimeters is indeed measured to ten percent accuracy, and a corresponding velocity is also desired to ten percent accuracy, the velocity error of 1.6 centimeters per second combined with a five minute averaging period yields an error of just over five milligals. One study of the filtering problem has been made by Moritz[17]. A study of the manner in which both altimeter and gravimeter readings can be used to obtain optimum results in the determination of vertical motion, as well as of the gravity reading, would also be useful.

The vertical navigation system used in the KC-135 aircraft consisted of an APR-5 profile recorder or radar altimeter, plus a hypsometer and a barometric altitude hold for the aircraft. The radar altimeter was used over surfaces of known altitude to provide a reference point for the hypsometer, which was then used as an altitude indication system for normal flight. The changes in altitude due to variations in isobaric surface height are compensated by means of Henry's correction[18].

3. THE GRAVIMETRIC SYSTEM

3.1 The Gravimeter

Inertial navigation has been used in continuous or "cruise" navigation of aircraft and ships and in launch guidance of rocket vehicles. The long periods of operation in cruise navigation make high demands on gyro accuracy, while requirements for accelerometer accuracy are relatively modest. In rocket vehicles the wide range of acceleration to be measured makes severe demands on accelerometer performance while, at the same time, long term stability is required if frequent calibrations are to be avoided. Consequently, the accelerometers used in rocket vehicles are logical candidates for gravimetric use.

The different types of accelerometers available include the pendulous gyro type used in both the SATURN and the TITAN III space boosters. The instrument used in the latter, known as the 25 PIGA (pendulous integrating gyro accelerometer) was the model selected for the gravimetry project.

The instrument has the general form of an octagonal aluminum cylinder approximately 11 centimeters in diameter and 15 centimeters long, with a weight of about 4 kilograms. The pendulous gyro within the instrument is a floated integrating gyro of beryllium construction, with a pendulosity of 22.5 gram-centimeters, and in normal use, an angular momentum of 30,000 dyne-centimeter-seconds. The floatation is aided by magnetic suspension produced by the microsyns at each end of the gyro. The pendulosity and angular momentum values give the device a scale factor of about 0.75 radian per second per gravity, in its normal manner of use. The servo system used to drive the accelerometer makes use of a direct drive d.c. torquer, and has a natural frequency of about 150 radians per second. The carrier frequency for the microsyn signals is 1000 hertz. The angular motion of the gyro turntable within the PIGA is measured by an optisyn[19], an incremental discrete readout device with an output of 2048 pulses per revolution.

In order to minimize the effort, cost, and time required to achieve an operating gravimeter from an existing accelerometer, whenever possible, use was made of existing electronic circuitry associated with the mechanical instrument. In the case of a gyro accelerometer, this circuitry involves the electronics necessary to operate the gyro, the servo amplifier used to drive the accelerometer torque motor, and the means of detecting the accelerometer output. The output had to be recorded in a form suitable for gravimetry.

The housing which was designed to provide thermal control for the instrument, whose operating temperature is about 63°C, also incorporated a mumetal layer for magnetic shielding. Two versions of the housing were designed; the second one differed by being symmetrical, so that the PIGA input axis could be pointed either up or down. This made it possible to use the instrument in whichever of the two directions incorporated the minimum thermal sensitivity.

One major change was made in the use of the instrument. The accelerometer was originally designed for the broad range of accelerations encountered by a rocket vehicle; in the gravimetric application a total range of two gravities is more than adequate. Consequently the scale factor of the instrument was increased by reducing the angular momentum stored in the gyro wheel. This was accomplished by design of a 50 hertz square wave supply to drive the gyro rotor. This reduced the angular momentum to 3750 dyne-centimeter-seconds, and increased the scale factor to 6 radians per second per gravity. Thus one PIGA revolution took place in about 1.05 seconds at one gravity.

The increased angular motion of the rotating PIGA element correspondingly reduces the sensitivity both to angular motion of the PIGA base and to uncertainties in measurement of the relative motion of the base and the rotating member. The net effect includes reduction of the time necessary to resolve a measurement to a given level of accuracy.

The 1000 hertz microsyn supply and the complete PIGA servo loop were left unchanged in form, although the gain of the servo loop was adjusted somewhat to correspond to the new gyro values, by resetting the potentiometer. The readout circuitry involved in the optisyn was also left unchanged.

The instrument output was recorded in the form of elapsed time per revolution of the PIGA. A Beckman preset counter, operating in a mode in which it is possible to count continuously without data loss, was used to count one megahertz pulses. Flight data were generally recorded as the elapsed time for one or two PIGA revolutions, corresponding to sampling intervals of approximately 1.05 or 2.1 seconds. The preset counter was connected through an intercoupler to a punched tape unit recording both the elapsed time for the PIGA revolution and the time from the Astrodata master clock in the aircraft, both in binary coded decimal form. A printed paper tape also recorded the elapsed time; the visual readout is convenient both during calibration and for checks during flight. Figure 5 is a photograph of the airborne electronics console containing the accelerometer electronics and readout equipment and the electronics used to operate the stabilization platform on which the PIGA was mounted.

At a later date a higher degree of system integration and simplification was obtained by using the frequency reference of the Beckman counter as the frequency source of the PIGA electronics as well; this increased the frequency stability of the latter and reduced the complexity of the overall system.

The accelerometer in the mechanical housing which included magnetic shielding, thermal regulation, and gyro and optisyn preamplifiers weighed approximately seven kilograms. The second version of the housing was a cylinder approximately 15 centimeters in diameter and 19 centimeters long.

3.2 Stabilization System

With the selection of a gyro accelerometer for use as a gravimeter, the stabilization requirement included either azimuth stabilization for the instrument or heading indication, as well as the vertical stabilization required for any instrument. If the instrument were to be used with its original scale factor, an azimuth error rate of approximately 0.155 degrees per hour corresponded to one milligal error in gravimeter indication. If the accelerometer scale factor were to be increased, the error sensitivity would be correspondingly reduced. The requirement for vertical stabilization was described earlier.

Since the goal was gravimetry rather than the design of stable vertical elements, it was hoped to find an existing platform that would meet the requirements. The azimuth stabilization requirements would place performance in the range of inertial navigation systems rather than of aircraft heading references, unless the accelerometer scale factor were drastically increased. A review of existing systems showed no available inertial navigation system or heading reference which had enough room on the stable member for the gravimeter. This is to be expected, since such systems are normally built as compactly as possible.

Previous tests using stabilized gravimeters had employed camera stabilization mounts, in particular Aeroflex Laboratories model ART-25, a two-axis stable element, and a later version, the ART-57. These platforms were equipped with a vertical gyro and two pendulums for levelling and employed direct-drive torque motors. The platforms tracked the apparent vertical rather than the true vertical, since there was no Coriolis compensation. In addition, the platform showed a limit cycle oscillation of about five arc minutes amplitude, probably due to non-linearities in the pendulums, and affected by the high erection rates required to overcome gyro drift. This performance was unacceptable; however, the ART-57 platform itself was phsyically large enough to hold the gravimeter and was available. In addition, the direct drive torquers could permit satisfactory stabilization to the vertical if adequate gyros, accelerometers, and servo electronics were used. The gyro accelerometer could thus be stabilized to the vertical and, if heading were indicated to sufficient accuracy, the stabilization requirements could be met.

A review of three-axis stable platforms was made in the hope that the azimuth axis and stable member of such a platform could provide the necessary attitude reference for the ART-57 gimbals. One suitable platform was the AN/AJN-10, an inertial reference platform built by AC Electronics Division, General Motors Corporation, for use in a Doppler-inertial aircraft navigation system. This platform, together with the associated AN/ASN-32 navigation computer and control indicator was available, and the aircraft was already equipped with a Doppler radar. Furthermore, the complete three-axis platform could fit on the stable member of the ART-57 platform, and the AN/AJN-10 platform also used direct-drive d.c. torquers. This suggested that integration of the two platforms might be fairly simple.

The AN/AJN-10 platform uses three single-degree-of-freedom floated integrating gyros and two floated restrained pendulums. The North and azimuth gyros are Honeywell GG49 MIG units; the East gyro, used for azimuth alignment in ground erection and for gyrocompassing in flight, is a 2FBG-2C model manufactured by AC Electronics. The pendulums are Honeywell GG59 units. The electronics are completely solid state, and the modes of operation are designed for convenient use. The platform weighs about 32 kilograms without electronics.

The complete AN/AJN-10 platform was placed on the ART-57 stable member, and the two outer axes, roll and pitch, of the AN/AJN-10 platform were locked. The servo signals which normally drove the torque motors on these axes were used instead to drive the torque motors of the ART-57 platform. The changes in the stabilization servo performance function needed to obtain satisfactory operation with the new moments of inertia and new torquers were achieved by the insertion of one resistor and one capacitor for each axis; even the power stage electronics of the AN/AJN-10 system was found adequate to drive the ART-57 torques. With this configuration, the internal changes in the AN/AJN-10 were limited to the addition of the two resistor-capacitor networks, and all other changes were accomplished in the external cabling. This permitted procedures of maintenance and calibration to remain largely unchanged.

The clamping mechanism used to lock the pitch and roll axes of the AN/AJN-10 platform was made adjustable, so that it could be used to align the stable element of the platform perpendicular to the input axis of the PIGA. The net result was a PIGA stabilized to the true vertical rather than the apparent vertical. The synchro signal on the azimuth axis of the AN/AJN-10 platform was available to indicate the heading changes of the PIGA base.

Laboratory operation of the combined platform indicated that the heading measurement errors would be too large if the scale factor of the PIGA were left unchanged. This was due in part to the fact that the AN/AJN-10 is a gyrocompassing system in normal airborne operation, and the high gain of a gyrocompassing loop leads to oscillations which take some time to damp out. Accordingly, the PIGA scale factor was changed as indicated earlier.

Finally, modification of the circuitry used to indicate the output of the azimuth synchro of the AN/AJN-10 was necessary to meet the desired resolution. This modification included an improved excitation to the synchro, and the use of a synchro control transformer, reset to null with the control transmitter on the gimbal at the start of each run. This arrangement made use of the fact that the anticipated gravimetric flights were all East-West or North-South rhumb-line courses, with anticipated heading changes within plus or minus ten degrees. Over this range the linearity of the synchro permitted an accuracy of ten arc minutes in heading indication.

Figure 6 is a photograph of the ART-57 gimbals with both the 25 PIGA and the AN/AJN-10 platform mounted on them.

4. TEST RESULTS

The flight test program was approved in April 1966, and the design and procurement of the necessary equipment started. The integration and modification of the stabilization system and of the gravity meter was accomplished by the end of September 1966, and aircraft installation followed. By late October the aircraft installation was undergoing the first flight checks, and in late November the aircraft proceeded to Carswell Air Force Base, Fort Worth, Texas, for initial flight tests over the designated gravity range area. The flight patterns followed parallels of latitude and meridians in the general area between 34° and 40°N latitude and 95° and 105°W longitude.

The AN/AJN-10 platform and AN/ASN-32 navigation computer had been designed for use with a velocity saturation level of 400 knots ground speed. It had been agreed to maintain the KC-135 aircraft speed below this level, but the November flights proved it impossible to keep aircraft speed within the specified limits at the designated altitude of about 7600 meters. Consequently, the aircraft returned to its base at Hanscom Field, Bedford, Massachusetts, and the AN/AJN-10 system was modified to raise the saturation level to 500 knots. The modification was essential because exceeding the saturation level caused error in platform verticality to accumulate at the rate of one degree in less than five minutes.

The aircraft then returned to Carswell, and seven days of gravimetric flights were conducted between December 11 and December 21, 1966, after which it returned once again to Hanscom.

In mid-January 1967 the aircraft had to be released for use in another project, but the schedule permitted one more return to Carswell and a few days of flight were possible before the aircraft was released. Since that time no aircraft has been available, but a C-130 aircraft was expected to be available in September 1967 for another series of flight tests.

Attempts were made to monitor the quality of stabilization system performance during these tests, both by comparison of the platform heading indication with that of the Astrotracker, and by observing the platform motion in levelling and in azimuth when it was restored to the ground erect mode after landing. Unfortunately, the power transfer from aircraft to ground power usually was sufficiently erratic to cause the AN/AJN-10 platform to downmode and lose the coordinate system it held. Figure 7 is the plot of the heading indication shown by the system on the one occasion when successful uninterrupted transfer to ground power was achieved; it indicates a change in azimuth of 20 arc minutes. Some of the comparisons of the astrocompass heading with the AN/AJN-10 heading indicated the possibility of azimuth performance problems in airborne operation.

After the December flight tests a switch was installed which permitted airborne operation using the Doppler system to damp the platform off-level oscillations, while the gyrocompass loop was open, so that the azimuth reference was held by the azimuth gyro alone. This switch permitted detection of cabling problems which introduced an azimuth error rate of about 1.5 degrees per hour, corresponding to about 1.2 milligals, during turns when the system was in pure inertial operation. The accumulated error was then offset by the gyrocompass loop which came into operation when the system returned to Doppler mode after completing the turn.

For the C-130 aircraft installation, it will be possible also to observe and record the AN/AJN-10 accelerometer signals, thus permitting some observation of in-flight platform verticality. In addition, switches have been included to permit removing Coriolis acceleration compensation, so that, if desired, the platform can operate in the "local vertical minus Coriolis acceleration" coordinate system described earlier.

Aircraft heading, as indicated by the azimuth synchro of the AN/AJN-10 platform, is shown in Figure 8, which represents typical flight conditions. The first half of the analog recording shows a period of moderate turbulence, with peak-to-peak oscillations of the order of one half degree. The second half shows quiet flight with peak-to-peak oscillations of the order of five arc minutes. The Dutch roll motion is evident throughout the recording, with a period of slightly over four seconds; the amplitude varies from about one arc minute peak-to-peak in quiet periods to about eight arc minutes in turbulence.

Some of the accelerometer results are presented in the next group of figures. Figure 9 is a record of the PIGA elapsed time, recorded for two- and ten- turns intervals, during ground calibration in the aircraft. Ground calibration was conducted before each flight, and the PIGA was then left in continuous operation until the end of the flight. Post-flight calibrations were also intended, but the power transfer problems discussed earlier usually interrupted PIGA operation. The scale of the recordings, in acceleration units, is the same, and the ten-turn records are approximately one fifth the amplitude of the two-turn recordings. Thus the variations apparently decrease inversely with the averaging interval in this region; a similar observation holds true for one-turn recordings. If the observations were considered as velocity readings obtained by integration over the appropriate period, the one-sigma value of the velocity readings would be the same for two- and ten-turn intervals. This indicates that most of the velocity power occurs at frequencies higher than the sampling rate of approximately one reading every 2.1 seconds.

Figure 10 represents an approximate spectrum of the ground test data obtained from the PIGA during an overnight run in which elapsed time for ten turns was recorded, using an approximate method of estimating spectra described by Blackman and Tukey[20]. The spectrum shows a peak in the vicinity of 0.0015 c/s, which represents the oscillation of the thermal control system of the instrument. In this particular case, inverting the instrument would have led to elimination of the peak, since the thermal sensitivity in the inverted direction was considerably lower.

Some flight records of PIGA output are presented in Figure 11, both for periods of turbulence and for quiet conditions. In quiet conditions, the one-sigma acceleration averaged over one minute of time with a two-second sampling interval, was generally about 2000 milligals. Only twice in all the flights did a value as low as 1000 milligals appear. This illustrates the level of vertical acceleration to be expected in KC-135 flight in the smoothest conditions obtainable.

The trace for intervals of turbulence shows peak-to-peak oscillations up to 140,000 milligals in amplitude, again with each reading representing a two-second average. Corresponding one-sigma values, averaged over one minute of time, showed a maximum value of 40,000 milligals. In more modest turbulence, frequently encountered, one-sigma values of 5000 milligals were common. The smallest observed one-sigma for a one-hour flight was about 3000 milligals; for a rough flight, about 10,000 milligals was observed.

Figure 12 shows an autocorrection function of the PIGA data obtained from a one-hour flight in smooth conditions, with the data points consisting of elapsed time for two PIGA turns, and the corresponding spectral density. The spectrum seems to be relatively flat between 0.01 and 0.10 c/s and then decreases slowly up to the limit imposed by the sampling frequency. The one-sigma value of the data was 3600 milligals for this flight.

The results of a one-hour flight in more turbulent conditions, with an overall one-sigma of 9400 milligals, are presented in Figure 13, again in the form of an autocorrelation function and spectral density. Here the spectrum peaks rather sharply between 0.04 and 0.05 c/s, this represents the performance function of the aircraft altitude hold system which is responding strongly to the turbulent air input.

During much of this flight, the conventional gravimeters were not in operation, because the accelerations exceeded the values at which the instruments would sustain damage. This demonstrated one of the major purposes of the flight test series, which was to show that the PIGA could continue to measure with satisfactory accuracy during atmospheric turbulence.

The gravimeter readings for test flights over known areas are compared with computed values of gravity at altitude, using Stokes's formula to compute the values at altitude from the observed gravity readings. However, the results presented in Figures 12 and 13 show the large amount of noise, due to vertical acceleration, present in the gravimeter readings. This once again illustrates the importance of a very high resolution barometric altimeter in obtaining accurate gravity observations or, alternatively, the limited resolution obtainable by filtering methods.

5. CONCLUSIONS

The flight test program has demonstrated the ready adaptability of inertial navigation equipment to the measurement of gravity aboard aircraft. It has also demonstrated some of the limitations of airborne gravimetry imposed by the lack of a suitably high resolution altimeter. Other inertial accelerometers are being tested in gravimetric experiments on surface ships, and they also demonstrate the suitability of such sensors to measurement in this environment.

In general, inertial instrumentation offers the possibility of holding angular coordinate systems with high precision, and of measuring linear motions aboard moving craft. The measurement of very small angles and of very small accelerations is also possible. It is quite possible that such instrumentation can be used to measure the behavior of the atmosphere and of the oceans, as well as for the measurement of the earth's gravitational field, and it is to be hoped that such applications will increase. It is also quite possible that the techniques used in building inertial gyros and accelerometers may also be applied to improve the design of instruments used in meteorology, oceanography, seismometry, and geophysics in general

ACKNOWLEDGEMENTS

The work described in this paper was supported by the United States Air Force Cambridge Research Laboratories. Technical supervision of the contract work was provided by Mr Bela Szabo and Mr Owen W.Williams. The senior author also wishes to thank Prof.Winston R.Markey of the Massachusetts Institute of Technology for interesting him in the problem of airborne gravimetry at a time when Prof.Markey was Chief Scientist of the United States Air Force.

REFERENCES

1. Heiskanen, W.A. Vening Meinesz, F.A. — *The Earth and Its Gravity Field.* McGraw-Hill, New York, 1958, pp.84-121.
2. Vening Meinesz, F.A. — *Theory and Practice of Pendulum Observations at Sea.* Pub. Ned. Geod. Comm., Delft, Waltman, 1929.
3. Worzel, J.L. Graf, A. — Bull. Geod., Vol.45, 1957, pp.38-52
4. Thompson, L.G.D. LaCoste, L.J.B. — Journal of Geophysical Research, Vol.65(1), Jan. 1960, pp.305-322.
5. Nettleton, L.L. et al. — Geophysics, Vol.XXV(1), Feb. 1960, pp.181-202.
6. Nettleton, L.L. et al. — Journal of Geophysical Research, Vol.67(11), Oct. 1962, pp.4395-4410.
7. Browne, B.C. — Mon. Not. Roy. Astr. Soc., Geophys Suppl., Vol.4, Sept. 1937, pp.271-279.
8. Dehlinger, P. Yungul, S.H. — Journal of Geophysical Research, Vol.67(11), Oct. 1962, pp.4389-4394.

9. Gantar, C. et al. Journal of Geophysical Research, Vol.67(11), Oct. 1962, pp.4411-4419.

10. Allan, T.D. et al. Journal of Geophysical Research, Vol.67(13), Dec. 1962, pp.5157-5162.

11. Thompson, L.G.D. Journal of Geophysical Research, Vol.70(22), Nov.15, 1962, pp.5599-5613.

12. Wall, R.E. et al. Journal of Geophysical Research, Vol.71(2), Jan.15, 1966, pp.465-485.

13. Talwani, M. et al. Journal of Geophysical Research, Vol.71(8), April 15, 1966, pp.2079-2090.

14. Jacoby, H. Schulze, R. Journal of Geophysical Research, Vol.72(8), April 15, 1967, pp.2199-2207.

15. Harlan, R.B. *Stabilization Requirements for an Airborne Gravimeter.* Massachusetts Institute of Technology, Experimental Astronomy Laboratory, Report RE-24, June 1966, pp.7-8, 39-41.

16. Harlan, R.B. *Eötvös Corrections for Airborne Gravimetry.* Massachusetts Institute of Technology, Experimental Astronomy Laboratory, Report RN-20, June 1967.

17. Moritz, H. *Optimum Smoothing of Aerial Gravimetry Measurements.* Ohio State University Research Foundation. Report 81, March 1967.

18. Henry, T.J.G. *Determination of Topographic Profiles by Use of Radar and Pressure Altimeters.* Meteorological Division, Department of Transport, Canada.

19. - *Digital-Analog Hybrid Shaft-Angle Indicator with Frictionless Optical Gearing.* Electromech. Des., 6, May 1958, p.12.

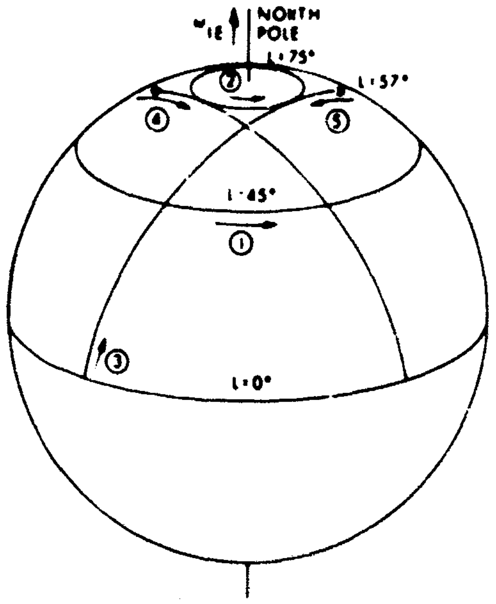
20. Blackman, R.B. Tukey, J.W. *The Measurement of Power Spectra.* Dover, New York, 1959, pp.42-47.

TABLE I

Vertical Platform Error Sensitivities

	Error Sources									
	A_{xo}	A_{yo}	A_{zo}	V_{xo}	V_{yo}	$(u)\omega_x$	$(u)\omega_y$	$(u)\omega_z$	$(u)G_x$	$(u)G_y$
A_x	$1\ \frac{\widehat{\text{min}}}{\widehat{\text{min}}}$	$1\ \frac{\widehat{\text{min}}}{\widehat{\text{min}}}$	$0.1\ \frac{\widehat{\text{min}}}{\widehat{\text{min}}}$	$0.22\ \frac{\widehat{\text{min}}}{\text{kt}}$	$0.22\ \frac{\widehat{\text{min}}}{\text{kt}}$	$0.2\ \frac{\widehat{\text{min}}}{\text{meru}}$	$0.2\ \frac{\widehat{\text{min}}}{\text{meru}}$	$0.02\ \frac{\widehat{\text{min}}}{\text{meru}}$	$0.0069\ \frac{\widehat{\text{min}}}{\mu g}$	$0.0034\ \frac{\widehat{\text{min}}}{\mu g}$
A_y	$1\ \frac{\widehat{\text{min}}}{\widehat{\text{min}}}$	$1\ \frac{\widehat{\text{min}}}{\widehat{\text{min}}}$	$0.1\ \frac{\widehat{\text{min}}}{\widehat{\text{min}}}$	$0.22\ \frac{\widehat{\text{min}}}{\text{kt}}$	$0.22\ \frac{\widehat{\text{min}}}{\text{kt}}$	$0.2\ \frac{\widehat{\text{min}}}{\text{meru}}$	$0.2\ \frac{\widehat{\text{min}}}{\text{meru}}$	$0.03\ \frac{\widehat{\text{min}}}{\text{meru}}$	$0.0034\ \frac{\widehat{\text{min}}}{\mu g}$	$0.0069\ \frac{\widehat{\text{min}}}{\mu g}$
A_z	$19.4\ \frac{\text{meru}}{\widehat{\text{min}}}$	$19.5\ \frac{\text{meru}}{\widehat{\text{min}}}$	$2.9\ \frac{\text{meru}}{\widehat{\text{min}}}$	$4.0\ \frac{\text{meru}}{\text{kt}}$	$4.1\ \frac{\text{meru}}{\text{kt}}$	$7.3\ \frac{\text{meru}}{\text{meru}}$	$7.5\ \frac{\text{meru}}{\text{meru}}$	$10.9\ \frac{\text{meru}}{\text{meru}}$	$0.063\ \frac{\text{meru}}{\mu g}$	$0.063\ \frac{\text{meru}}{\mu g}$

These sensitivities represent peak errors resulting from initial platform misalignments, initial velocity uncertainties, and steps in gyro drifts and accelerometer biases.



① 45° latitude course, 300 knots east
② 75° latitude course, 300 knots east
③ Great circle course, 0° to 75° to 57° latitude, 300 knots
④ Great circle course, 57° to 75° to 0° latitude, 300 knots
⑤ Great circle course, 57° to 75° to 0° latitude, 300 knots

Fig. 1 Navigation courses for Schuler system study

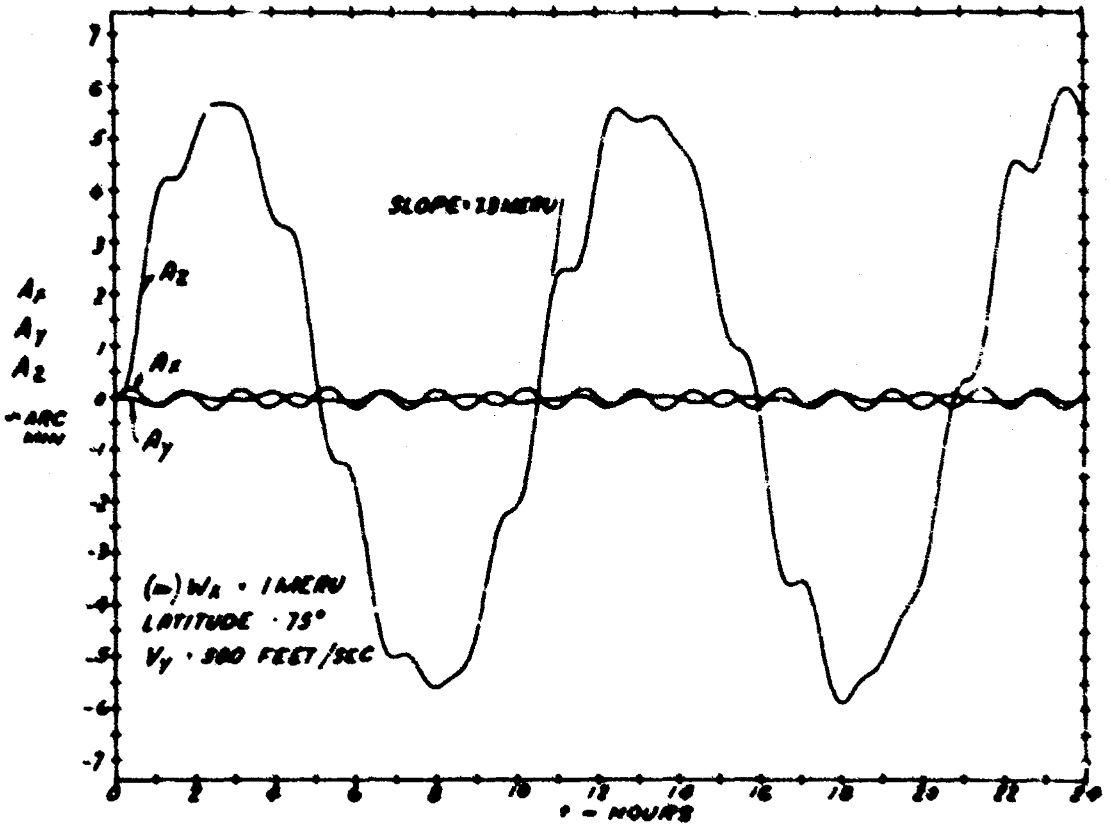


Fig. 2 Platform angle responses to North gyro drift

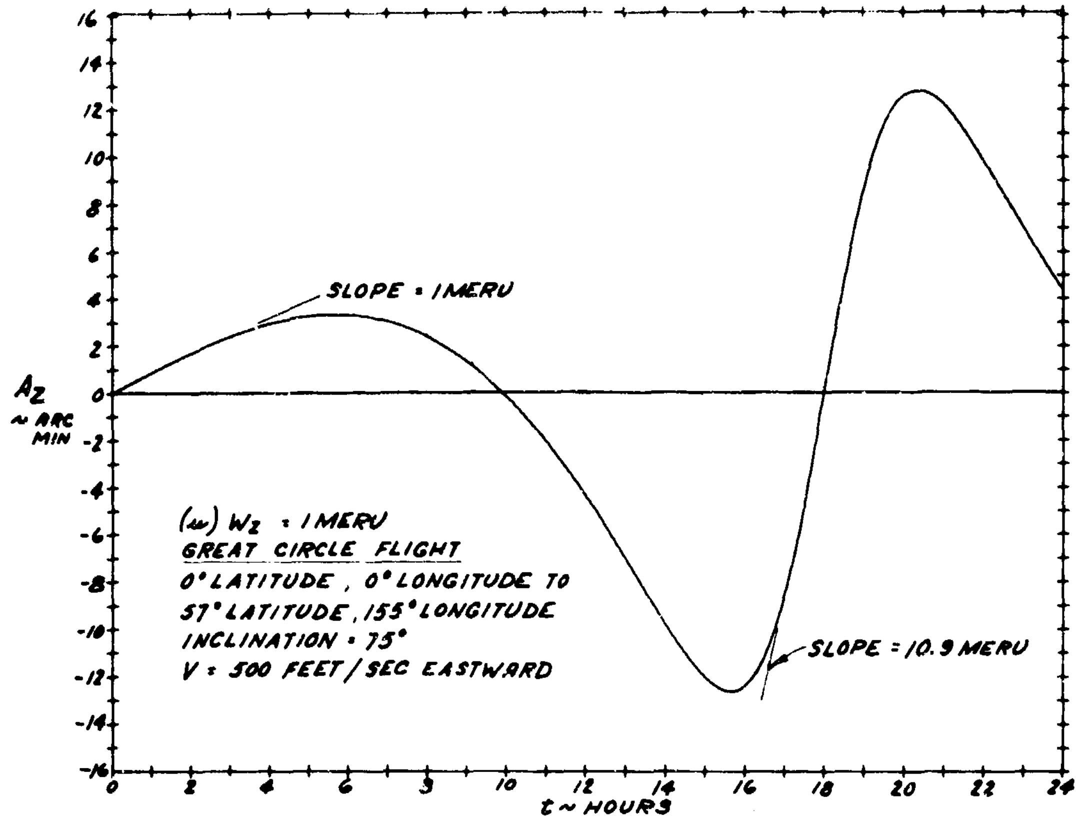


Fig.3 Platform azimuth error due to azimuth gyro drift

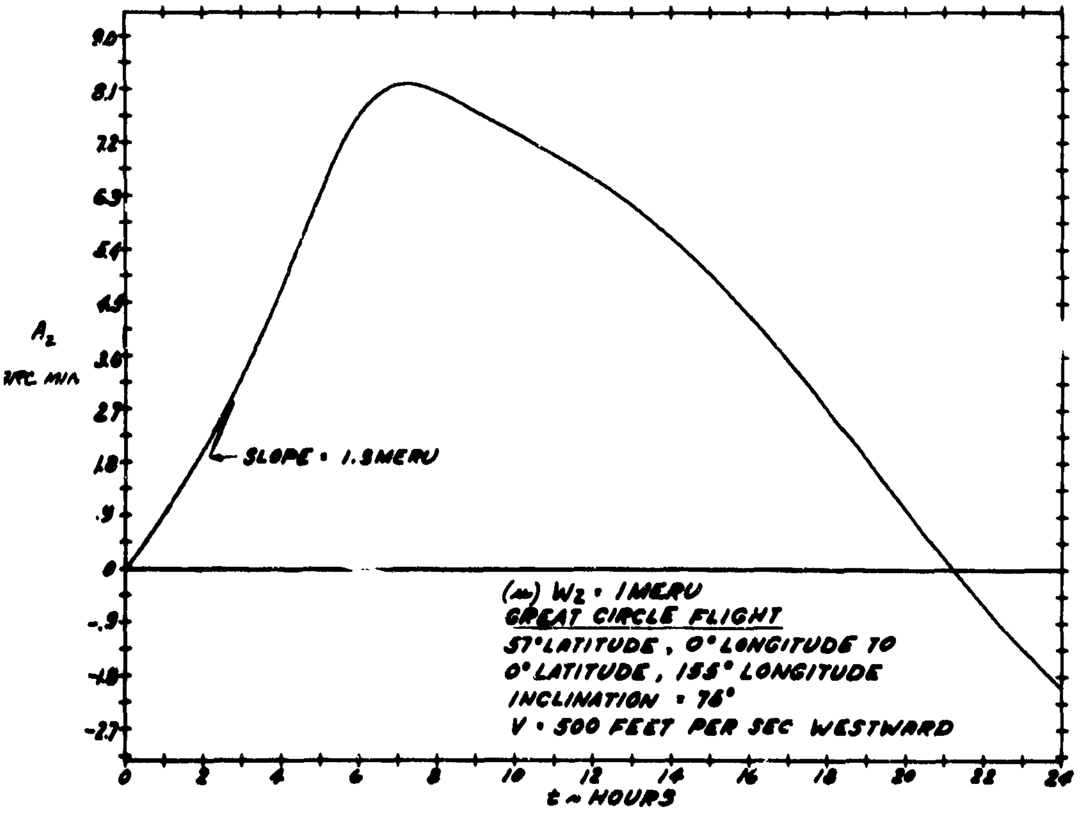


Fig.4 Platform azimuth error due to azimuth gyro drift

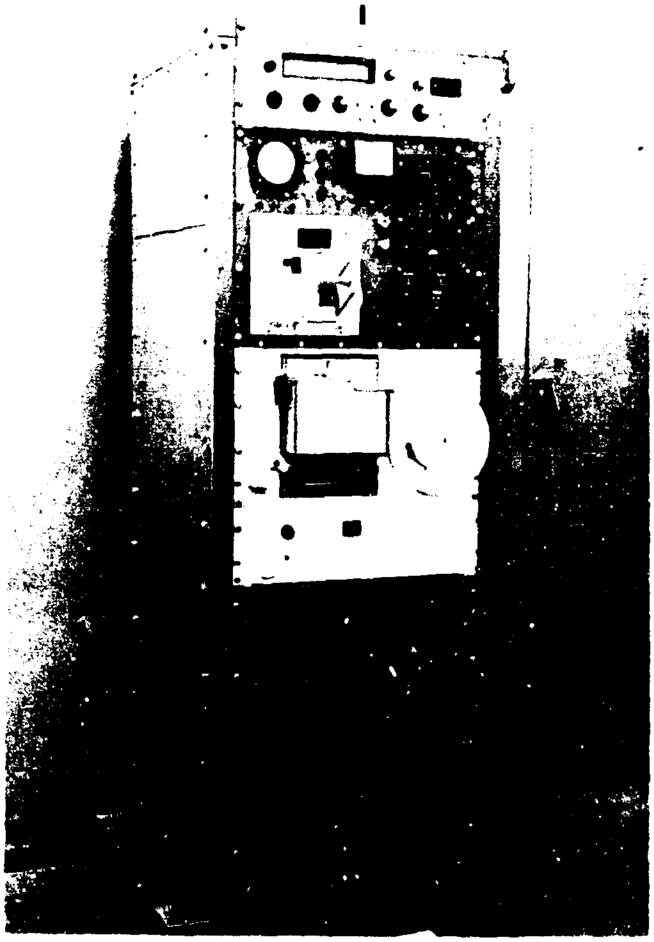


Fig.5 Airborne electronics console

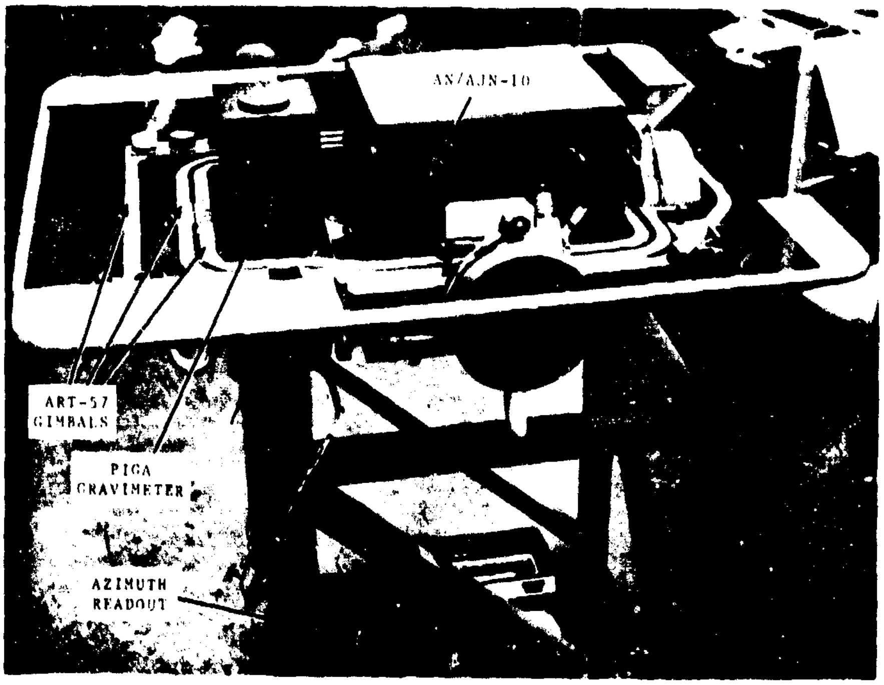


Fig.6 25 PIGA gravimeter on stabilized platform

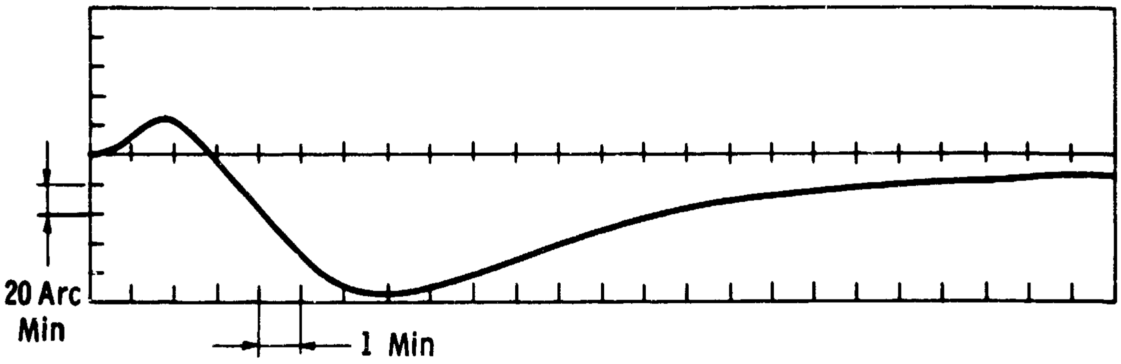


Fig.7 AN/AJN-10 azimuth performance as shown by post-flight gyrocompassing

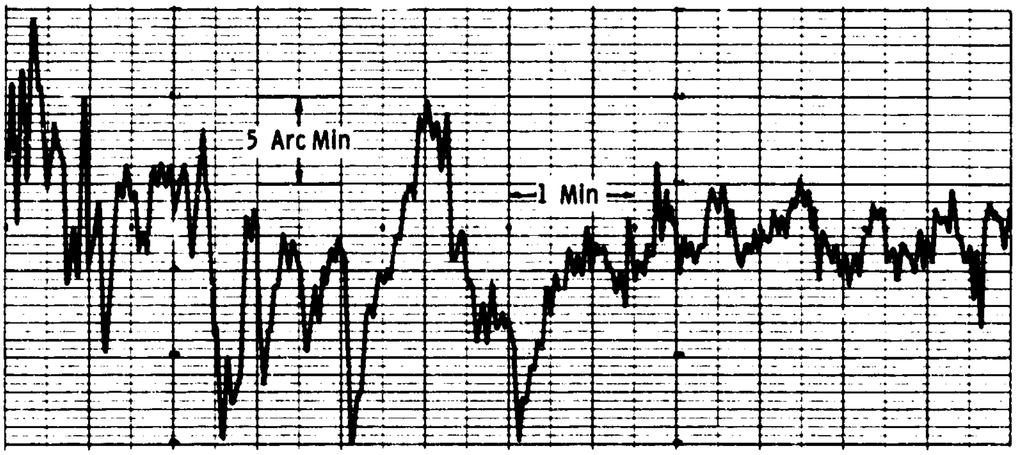


Fig.8 Aircraft heading indication from AN/AJN-10 platform

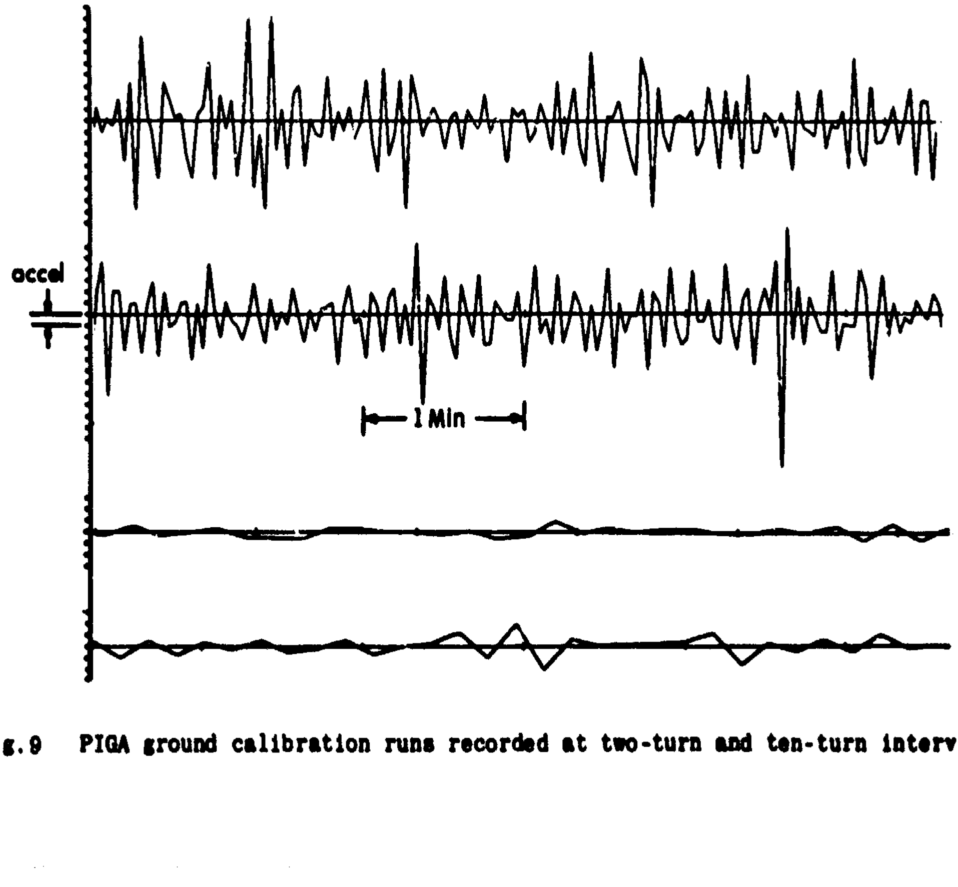


Fig.9 PIGA ground calibration runs recorded at two-turn and ten-turn intervals

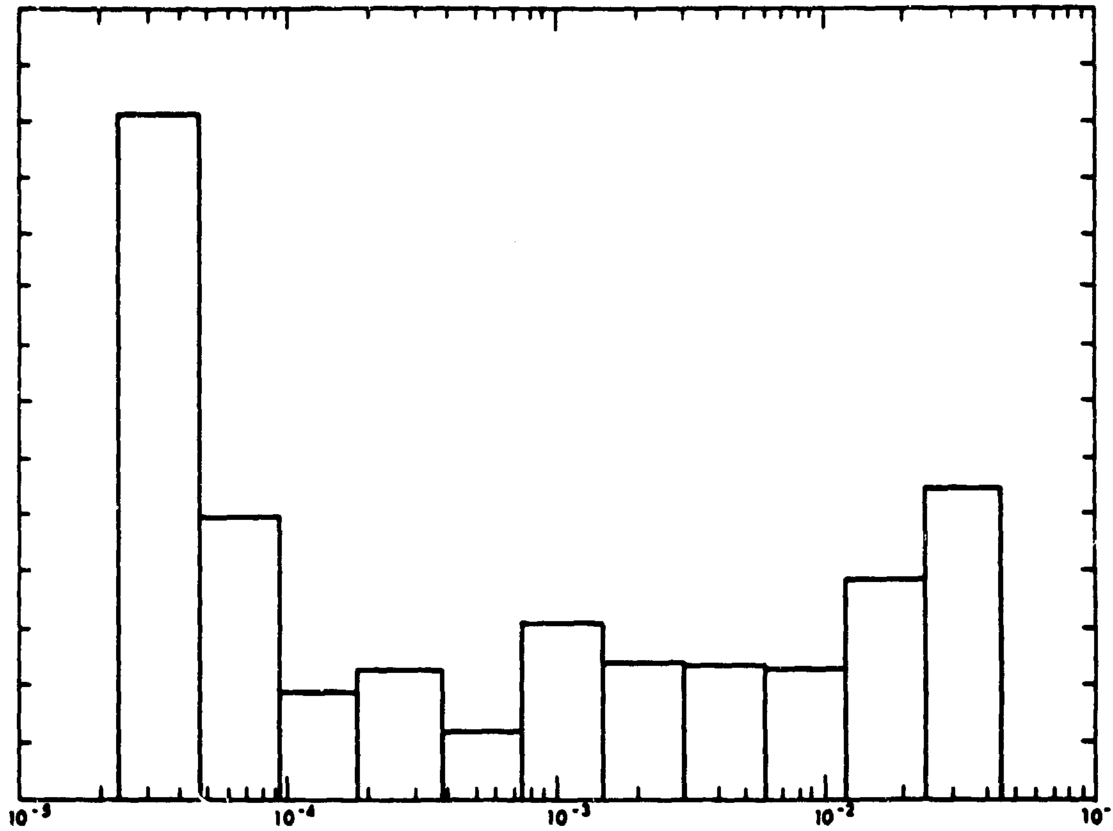


Fig.10 Pilot-estimated spectrum of PIGA overnight ground test. Ten-turn recordings

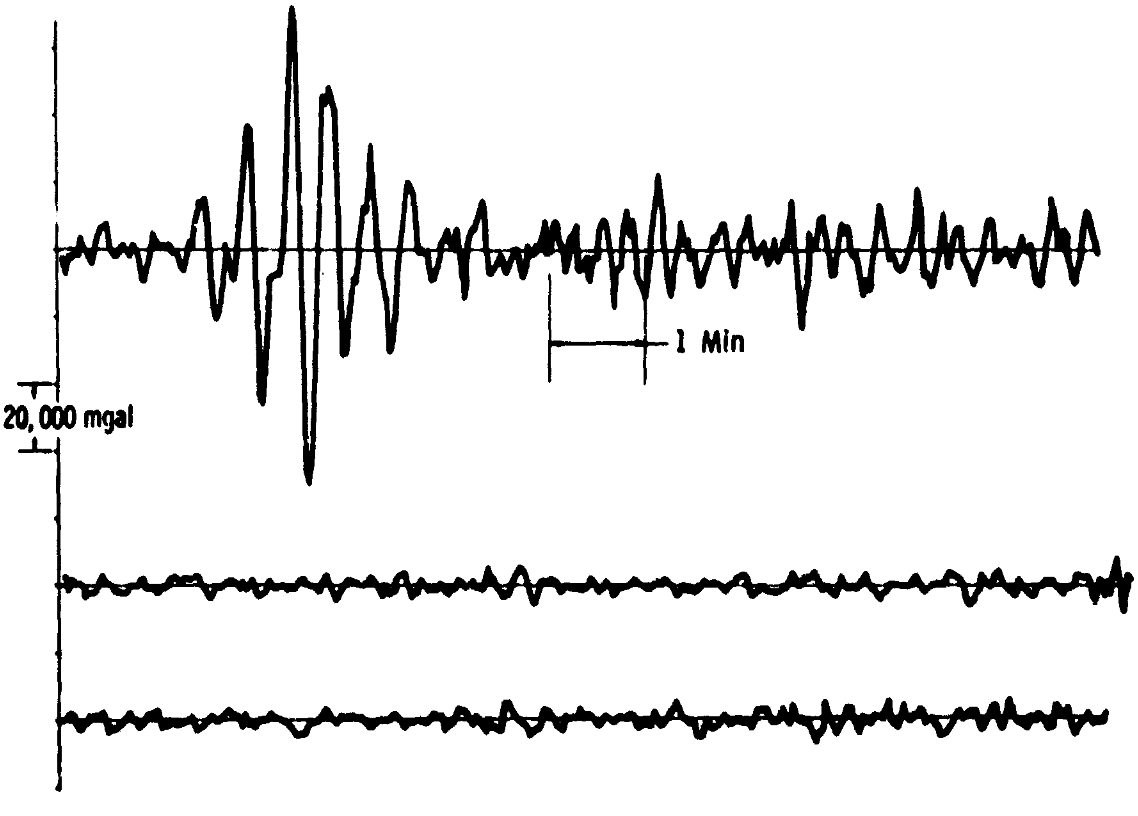


Fig.11 Aircraft vertical acceleration as indicated by two-turn PIGA readings, for turbulent and quiet conditions

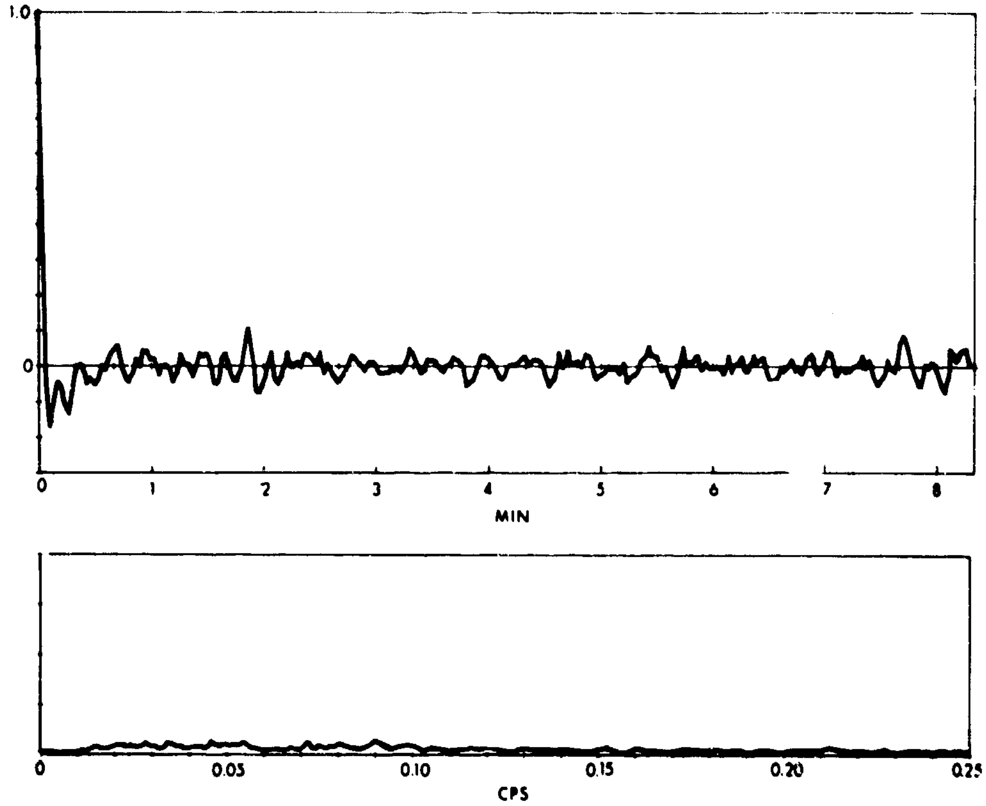


Fig. 12 Autocorrelation function and power spectral density of vertical acceleration in smooth flight

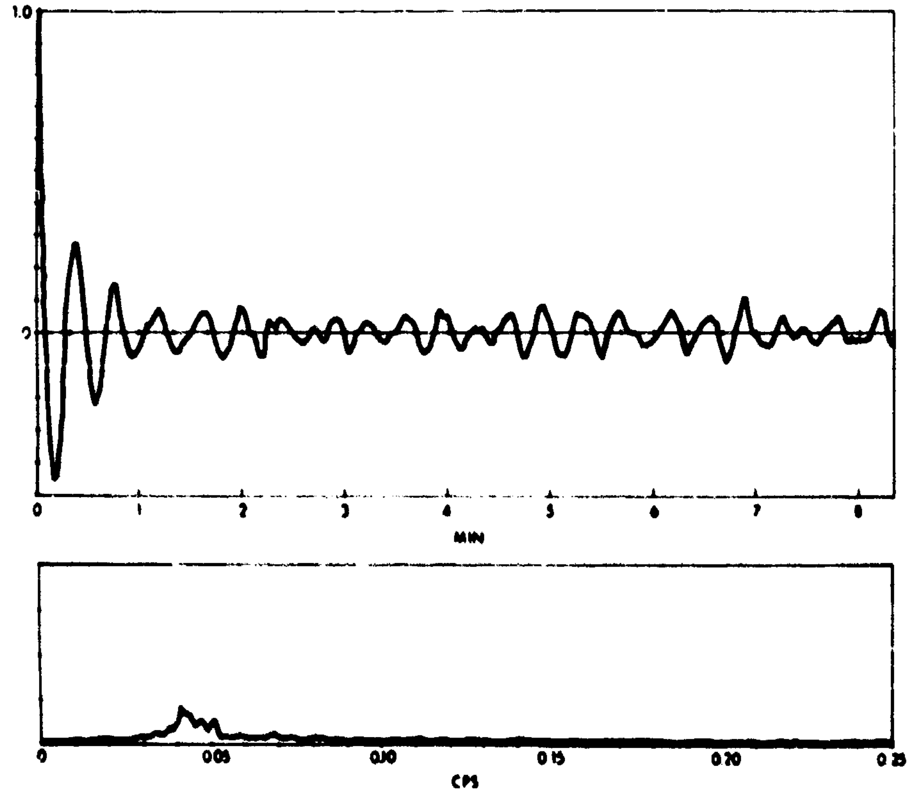


Fig. 13 Autocorrelation function and power spectral density of vertical acceleration in turbulent conditions

PRESENT-DAY NORTH-FINDING SYSTEMS AND THEIR FUNDAMENTAL LIMITATIONS

by

B. T. Trayner

British Aircraft Corporation,
Stevenage, Herts, England

SUMMARY

In the past few years portable gyroscopic North-finders for use on land have evolved which have errors as low as a few seconds of arc. These have mostly been made to meet military needs for survey and gun-laying and civil needs for underground survey. Three basically different types of north-finder have been used in commercial systems which use respectively a pendulous gyro, a single-axis rate gyro and a two-axis rate gyro. The basic principles and method of use of each system are described and the sources of error including those which arise from the environment in which the system is used are outlined.

It is indicated that present day systems are limited in accuracy by the environment in which they are used and by the time available as well as by the usual gyro errors. They are limited in their application by their cost and this is the field where most development is likely in the immediate future.

NOTATION

D	gyro drift (of single-axis gyro with output axis vertical)
g	acceleration due to gravity
H	angular momentum of gyro rotor
M	mass of gimbal and wheel
l	distance between centre of gravity and suspension point of gimbal (of gyro-pendulum)
R	component of earth's rotational rate in the horizontal plane at an angle θ from North
R_I	rate indicated by a single-axis gyro when measuring R
R'	component of earth's rotational rate in the horizontal plane at an angle $\theta + \phi$ from North
R_I'	rate indicated by a single-axis gyro when measuring R'
α	angle between sphere and case axes of rotation (two-axis rate gyro)
α_0	equilibrium value of α
γ	period of oscillation about the meridian plane (of gyro-pendulum)
λ	latitude
θ, ϕ	angles, see definition of R and R'
ψ	angle between plane defined by sphere and case axes of rotation and the meridian plane (two-axis rate gyro)
Ω	earth's rotational rate, equal to 15.041 degrees per hour

PRESENT-DAY NORTH-FINDING SYSTEMS AND THEIR FUNDAMENTAL LIMITATIONS

B. T. Trayner

1. INTRODUCTION

Over the past decade, advantage has been taken of the techniques exploited for inertial navigation to produce portable, light-weight gyrocompasses intended for surveying, especially underground and for military purposes. These devices have all been designed for operation on a stable, stationary base, i.e. the ground, so the errors to which they are subject are rather different to those of the conventional ships gyrocompass, as also are the allowable error magnitudes. The accuracy demanded ranges from a few minutes to a few seconds of arc. All gyrocompasses rely on detecting the direction of the horizontal component of the earth's rotational rate. To place the performance required from the gyrocompass in perspective with other inertial navigation gyros. Figure 1 illustrates the rates which the gyrocompass must be capable of detecting to achieve any given North-finding accuracy. This is of course dependent on the latitude.

Three basic types of North-finder have been used in commercial devices. These may be termed the gyro-pendulum, the single-axis rate gyro and the two-axis rate gyro. Their principles, basic construction and method of use will be briefly described and the main sources of error and the relative advantages of the different systems will be discussed.

2. PRINCIPLES OF OPERATIONS

2.1 The Gyro Pendulous North-Finder

If a rotating wheel (see Figure 2) is suspended by a fine suspension wire with the spin axis approximately horizontal, the equilibrium position is when the spin axis lies in the meridian plane. If the system is only very lightly damped, then equilibrium is not reached for a considerable time after a disturbance, but an estimate of the equilibrium position can still be made. Such a pendulum has several modes of oscillation, mostly of short period, but the dominant, long period mode is one in which the spin axis describes (ignoring the damping) an elliptical cone, with the major axis of the ellipse horizontal. The period τ of this oscillation is given to a first order by (see Notation)

$$\tau = 2\pi\sqrt{\left(\frac{H}{Mgl\ \Omega\cos\lambda}\right)} \qquad (1)$$

The direction of North is the mean of the easterly and westerly excursions of the spin axis, allowance being made for the decrement due to the damping.

2.2 The Single-Axis Rate Gyro North-Finder

This system relies on making measurements of the horizontal component of the earth's rotation. The value of this rate R is given by

$$R = \Omega\cos\theta\cos\lambda .$$

To solve this equation for ψ without ambiguity a further measurement is required,

$$\text{i.e.} \qquad R' = \Omega \cos(\psi + \Phi) \cos\lambda ,$$

where Φ is a known angle.

In practice, to avoid precise measurement a value of Φ is found where $R = R'$, when it may be deduced that $\psi = 180^0 - (\Phi/2)$. This is illustrated in Figure 3(a), where R is plotted against the azimuth.

2.3 Two-Axis Rate Gyro North-Finders

If a two-axis rate gyro is mounted with its spin axis vertical, then, provided its sensitivity is the same to all inputs in the horizontal plane, the direction in which the spin axis is deflected will bear a direct relationship to the meridian. In the equilibrium position the gyroscopic torque due to the earth's rotation will be balanced by a torque due to the deflection of the spin axis.

3. GENERAL CONSTRUCTION AND METHOD OF USE

The North-finder must have some method of transferring its information to other equipment. The theodolite is the conventional instrument which has been generally used and incorporated with the North-finder. A theodolite needs to be a certain minimum height above the ground to give it a reasonable field of view and is used on a tripod. This practice has been followed with all the portable North-finding systems which, to the author's knowledge, are commercially available, and the gyroscopic unit is mounted either above or below the theodolite.

Some of the typical features of available North-finding systems are now given. It must be emphasized that these are not the only possible systems.

3.1 The Gyro-Pendulous North-Finder

A gyroscopic unit of the type already described (Fig.2) is attached to the azimuth axis of the theodolite. The wheel is run at constant speed by a synchronous motor. The suspension wire can be used as one lead for the motor but two other fine leads must be provided, which should have as little influence on the pendulum as possible. A pick-off, usually optical, is used to detect rotation of the gimbal in azimuth relative to the case of the unit. A method of clamping the gimbal is necessary to enable the equipment to be transported.

In use the suspension wire may be adjusted so that the equilibrium position of the gimbal coincides with the pick-off null when the wheel is not rotating. With the wheel at synchronous speed and the spin axis approximately North, the gimbal is unclamped, care being taken not to excite the unwanted modes of oscillation. As the gimbal swings the theodolite is rotated in azimuth to maintain the pick-off at its null. The turning points of the swing can be read from the theodolite azimuth circle. Three such turning points are required as a minimum to estimate the equilibrium direction of the spin axis. Turning the theodolite in step with the gimbal avoids twist in the suspension and enables the torque for this suspension to be kept zero. This may be done manually by the operator or, at the cost of greater complexity, automatically.

No attempt is normally made to mechanically align the spin axis, the pick-off and the theodolite telescope or circle, but the instrument is tested against a known direction and a correction factor is found. Ancillary equipment required in the field is an alternating current source of constant frequency to drive the spin motor.

3.2 The Single-Axis Rate Gyro North-Finder

The gyro which has been used with this type of North-finder is the single-degree-of-freedom rate integrated gyro similar to that commonly used in inertial navigation systems. The gyro is converted to a rate gyro by an external electronic circuit, which takes the output from the gimbal pick-off and feeds a current to the gyro torque motor proportional to the gimbal deflection. This gyro, with its output axis vertical, is attached to the azimuth axis of the theodolite. The gyro input axis is in the horizontal plane and the current in the torque motor is proportional (forgetting gyro errors) to the component of the earth's rotation in a particular horizontal direction.

In the field the method of use is as indicated in Section 2.2. Two directions are found in which the rates indicated by the gyro are the same, the directions where the rate is zero being chosen since these can be most accurately found. The fact that the rate indicated by the gyro differs from the actual rate by some quantity due to gyro drift is of no consequence provided this drift remains constant. If the gyro drift is D and the rate indicated by the gyro is shown by the suffix I, then

$$R_I = R + D$$

$$R_I' = R' + D$$

and if $R_I = R_I'$ then $R = R'$.

This is illustrated in Figure 3(b), where R_I is plotted against the azimuth of the gyro input axis.

It is not necessary to point the gyro input axis exactly to find the direction of zero indicated rate. Provided the input axis is close to the direction, a correction can be made with sufficient accuracy from the actual rate measured.

As with the gyro-pendulous North-finder, mechanical alignment of the gyro unit with the theodolite is difficult and the usual procedure is to calibrate.

Ancillary equipment required in the field consists of an alternating current source of constant frequency to drive the spin motor, a heater supply and temperature controller to maintain the gyroscope at constant temperature, and the electronic circuitry gyro between pick-off and torque motor already mentioned.

3.3 The Two-Axis Gyro North-Finder

The gyro which has been used in this type of North-finder is illustrated in Figure 3. A spherical rotor is supported on an aerostatic bearing within a case which is in turn mounted on a pair of bearings and rotated at high speed about a vertical axis. In the absence of any external rate the sphere and case would rotate together on a common axis, but the earth's rotation makes the sphere axis move in a westerly direction to an observer looking down on the gyro. As the two axes diverge a torque is generated between sphere and case, due to the viscous forces, which tends to precess the sphere axis back into coincidence with the case axis. In the equilibrium position the plane containing both the case and sphere axes makes an angle ψ to the meridian plane given by

$$\psi = 90^0 - \alpha_0 \tan \lambda ,$$

where α_0 is the equilibrium angle between case and spin axes.

A small mirror is mounted on both the case and the sphere, normal to the axes of rotation. Using an autocollimating eyepiece on the theodolite the attitude of both axes may be determined.

Ancillary equipment required consists of a power source for the drive motor and a compressor to supply air to the bearing.

4. ERRORS IN NORTH-FINDING

For convenience the errors which arise can be considered as originating in the gyro, the rest of the system or from the environment. Time is an important parameter with many of the errors, and time also affects the number of North determinations which can be made, which influences the effects of truly random errors. Time is all-important to the military user; the civil user is prepared to take longer if it gives a definite improvement in accuracy. The consequences of this will be discussed at greater length.

4.1 Gyro Errors

With the gyro-pendulous and single-degree-of-freedom rate gyro any gyro error arising from an unwanted torque acting on the gimbal can be allowed for, provided that it is fixed. Such torques may arise from the gimbal suspension, the gimbal pick-off, the method of conveying power to the gyro wheel and by the action of electric and magnetic fields (which may be generated by the gyro motor). With the single-axis rate gyro it is only essential that these torques remain constant during a measurement. Their absolute value does not affect the calibration. However the torques which act in the gyro-pendulum which cannot be checked whilst the wheel is not operating, i.e. any which arise from the motor field, can only be allowed for in the calibration, and any change will affect the calibration.

By virtue of its construction the only torques of those just listed likely to act on the sphere of the two-axis systems which has been described are those from the air bearing support, and these average to zero as the case is revolved.

A further source of error in most gyros is the relative position of the centre of gravity of the gimbal and the centre or axis of suspension. This is of no consequence to the single-axis gyro, since the axis of freedom is vertical. The pendulous-gyro depends for its action on a displacement between the two, and is not affected by minor changes in their relative position provided that they do not occur during a measurement. If such changes occur suddenly a change in the amplitude of the swing takes place, and readings before and after the change cannot be related. If the centre of gravity should move steadily (a "ramp" in inertial gyro terminology) then there will be a shift in the indicated North, but such ramps would have to be very large to have a significant effect. Any displacement of the centre of gravity of the sphere in the two-axis gyro normal to the spin axis is of no concern, since, as with the suspension torques, the resultant torque is averaged to zero. However, any displacement of the centre of gravity from the centre of suspension along the spin axis has an immediate effect when the spin axis moves from the vertical. This gives a deviation of the indicated North which must be allowed for by calibration.

There are other sources of gyro error which generally are of no concern to an inertial system but which adversely affect the North-finding system. These are errors which give an appreciable error over the time a North-finder must be used, but never integrate to a large enough value to affect a typical inertial system. One of these is the effect of movements of the spin axis within the gimbal or, in the case of the two-axis gyro, movement of the case rotation axis. These movements can and do appear, especially when the axis is defined by a pair of ball bearings, where they have a tendency to be periodic due to relative motions of balls and cages within the bearings. As an illustration of their possible effect, a 0.1 second of arc shift in the spin axis over 100 seconds of time may be considered as a drift of 0.001°/hour.

4.2 Other System Errors

The gyro unit in the case of the gyro-pendulous or single-axis North-finders as described is mechanically coupled to the theodolite vertical axis, and the calibration factor depends directly on the angular stability of this coupling. This does not affect the two-axis rate gyro where the theodolite is also used as the gyro pick-off, but a considerable error can arise with this system because of the read-out. It defines a plane by two lines (the two axes of rotation) which are typically of the order one degree to each other. This means that, to attain an accuracy of one arc minute in the north determination, the spatial attitude of the axes must be determined within one arc second. It is very difficult to increase the angle between the axes, since this means reducing the viscous drag between sphere and case, which can only be done at the cost of large clearances in the bearing, much higher gas flows and hence larger compressors.

4.3 Errors Due to Environment

In the Introduction it was stated that the North-finders discussed were for use on a stationary, stable base, but stable is only a relative term and the very small tilting motions which occur in the ground and in the support structure used to carry the system are critical to the accuracy of the North-finding system.

These tilts may be continuous or periodic. Continuous tilt can arise from differential expansion within the support structure (usually a tripod) due to temperature or humidity effects and from sinking of the support structure into the ground. Oscillatory tilts are always present in the ground and are greatest near moving vehicles, machinery etc. A further cause is wind-excited vibrations of the support structure and gyro-theodolite.

The gyro-pendulum is effectively isolated from the continuous tilts (provided it is not damped in any way) and they should have no effect on it. The two rate gyro systems cannot distinguish between such rotations and the polar rotation of the earth, and will therefore produce an error. These errors can be allowed for by measuring the rate and direction of tilt of the gyro-theodolite relative to the local vertical by some other method, though this has not generally been done.

The oscillatory tilts have two effects. They give a periodic component to the output of the single-axis rate gyro and an oscillatory indication of North from the two-axis rate gyro, but these frequencies are generally above 1 c/s. The other effect is the rectifying effect known as "coning". If the input axis of a rate gyro is forced to describe a cone the instrument sees an input rate which is equivalent to the solid angle of the cone in the time taken to describe it.

This is unlikely to affect the two-axis rate gyro because, at frequencies above about 0.01 c/s, the gyro behaves as a free gyro, i.e. the spin axis does not appreciably move from its spatial position, but the single-axis gyro will see the error in full. To illustrate its magnitude, if the input axis describes a cone of semi-angle 2 seconds of arc at 10 c/s, this is equivalent to an input rate of 0.0006°/hour.

Though the pendulous-gyro is effectively isolated from such oscillatory tilting of the gyro system, similar errors can arise due to lateral vibrations of the suspension point, especially if the vibration frequency lies close to one of the basic pendulous frequencies. Such effects could be reduced by damping the pendulum, but this would make the system sensitive to tilts.

5. THE INFLUENCE OF TIME ON ACCURACY

5.1 The Gyro-Pendulous North-Finder

Systematic errors such as are due to wrong calibration, unwanted gimbal torques etc. are not reduced by long time of testing, but those due to transient effects, random observer error etc. will reduce as the inverse of the square root of the number of turning points observed. Certain other errors due to the environment may be reduced by North-finding at several distinct points and correlating the results. (This is true of the other two types of North-finder as well). It is obvious that the feature which determines the number of readings which can be taken in any given time is the period (see Equation (1)) and from this aspect the shorter the period the better. However, we require a large value of H (to keep the effect of unwanted torques small) and a small value of M (to enable a delicate low torque suspension wire to be used). This leaves the length l as the only other feature which is under the designer's control. This is limited by the physical size of the system and the increase in M which is necessary if the gimbal is to be kept as a rigid structure. The majority of commercial North-finders have periods of between five and ten minutes.

5.2 The Single-Axis Rate Gyro

There are three factors to be considered, the time taken to obtain a reading, the time between successive readings, and the number of readings. The number of readings influences the probability of error from random causes, as with the gyro-pendulum.

The time between successive readings should be short to reduce the probability that the gyro fixed torques have changed, but the time taken over each reading must be long enough to allow the averaging or smoothing out of the transient rates which are always present, and time must also be allowed for certain systematic effects to disappear after the gyro has been aligned to a new azimuth. These effects are dependent on the method which has been used to define the output axis of the rate gyroscope. If the gimbal is supported on jewels and pivots, these must operate with clearance and the gimbal has a finite angular freedom about the input axis. Before an accurate rate reading can be made the output axis must have reached a stable position, which depends on the vertical component of the earth's rate. The time taken to reach this stable position is largely dependent on the damping of the gimbal about the input axis. This damping is very high with the conventional design of single-axis rate integrating gyroscope but it may be reduced by, for example, introducing channels into the damping gap.

5.3 The Two-Axis Rate Gyro

A North value may be taken with this system from the moment the system has been set up and the case and sphere run up to speed. When the sphere is decaged the axis of the sphere immediately commences to move off in the East-West vertical plane, and the accuracy increases as the divergence between case axis and sphere axis increases. The angle between these two axes α is given to a first order by

$$\alpha = \alpha_0(1 - e^{-t/t_0})$$

where α_0 is the final, equilibrium value of α and t_0 is a time constant which will have a value of the order five minutes if $\alpha_0 = 1^\circ$.

After the sphere axis has approached its equilibrium position, extra time can be used to take repeated observations in the theodolite, taking the usual step of reversing the horizontal theodolite axis between readings.

6. COMPARATIVE PERFORMANCE

Performance is, as always, largely a question of what one is willing to pay for, no matter what system is being considered. The following comments are, necessarily, very general.

There is very little published data on the two-axis gyro system; it has not been developed as extensively as the other two, and it is difficult to make a comparative assessment of its potential as its main problems are essentially peculiar to that system only.

Of the other two systems, the gyro-pendulum is relatively simple in construction but is not so simple to operate and the suspension is prone to damage. It can be made easier to operate by the addition of automatic read-out systems or by the inclusion of damping, but these increase the cost and complexity and the second feature can increase the liability to error. The single-axis rate gyro is both very rugged and easy to operate. The response of the two systems to their environment is different, i.e. the undamped pendulum is affected by horizontal vibrations, the single-axis gyro by tilt and the damped pendulum by both. There is not enough evidence to indicate which of these is likely to be the more severe problem, but the fundamental limitations of the simple systems which have been described are from the environment and not the mechanics of the system. To overcome these limitations it will be necessary to isolate from the environment or to measure the effects by other instruments and make a correction for them.

There is not a great deal of difference at this stage of development between the accuracies claimed by the different manufacturers; typically these are of the order one minute of arc half an hour after setting up, improving to ten seconds of arc after repeated readings.

7. FUTURE DEVELOPMENTS

North-finders are now available which will, in the field, meet most of the accuracy requirements of potential users. Future development must be aimed largely at reducing both the capital cost and the operating cost of the equipment. This development will be very much influenced by the parallel development of inertial navigation systems where the aim is similar. New types of angular rate or displacement sensors are reported regularly and some of these may be adaptable for use as part of a North-finding system. Past experience has shown, however, that the mechanical design of the system is only one factor. Developing a suitable operating technique is of equal importance to the overall accuracy, and, as emphasised in the previous section, the environment in which the system is used must be carefully considered.

At present we often calibrate our North-finders by comparison with a datum obtained from astral survey. In the next few years it is reasonable to expect that this will not be adequate and that the gyroscope will rival in accuracy the older methods of determining the meridian.

REFERENCES

For greater detail on the construction and method of use of specific systems the following may be consulted:

The Gyro-Pendulous North-Finder

Lauf, G.B.	*The Gyro Theodolite and its Application in the Mining Industry of South Africa.* Journal of the South African Institute of Mining and Metallurgy, March 1963, pp.349-386.

The Single-Axis Rate Gyro North-Finder

Thomas, T.L.	*Precision Indicator of the Meridian.* Chartered Surveyor, March 1965, pp.492-500.
Thomas, T.L.	*The Precision Indicator of the Meridian, Theory and Practice.* Proceedings of the Third South African National Survey Conference, Johannesburg, January 1967.

The Two-Axis Rate Gyro North-Finder

	UK Patent No. 1,016,260.
Klass, P.J.	*North Seeking Gyro Proves Fast, Accurate.* Aviation Week and Space Technology, December 17, 1962.

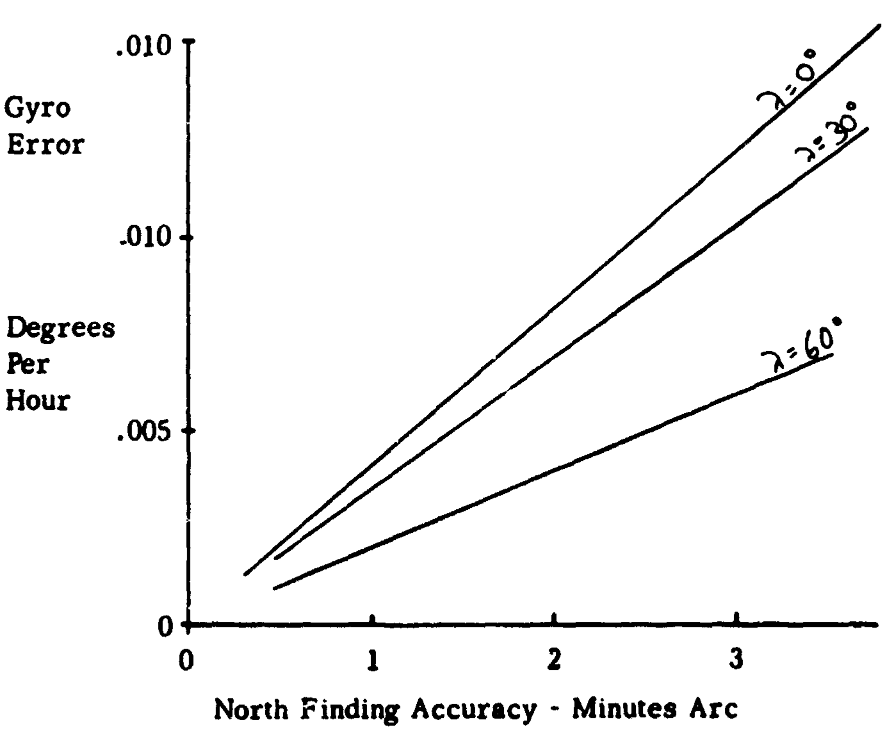


Fig. 1 Relationship between gyro and North-finding accuracies

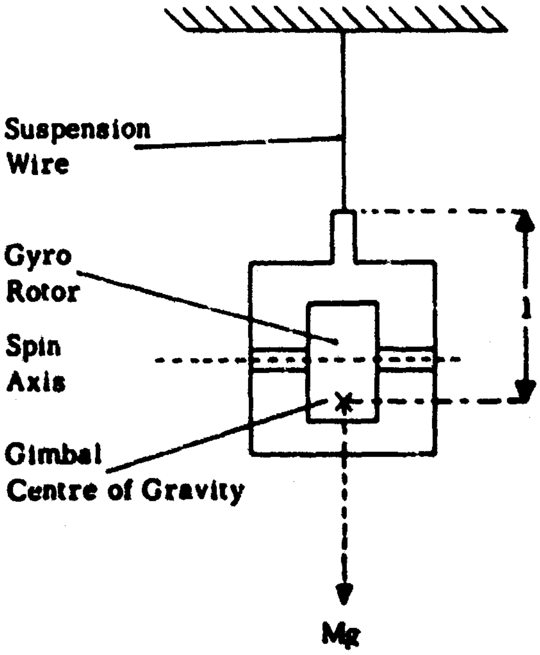


Fig. 2 The gyro-pendulous North-finder

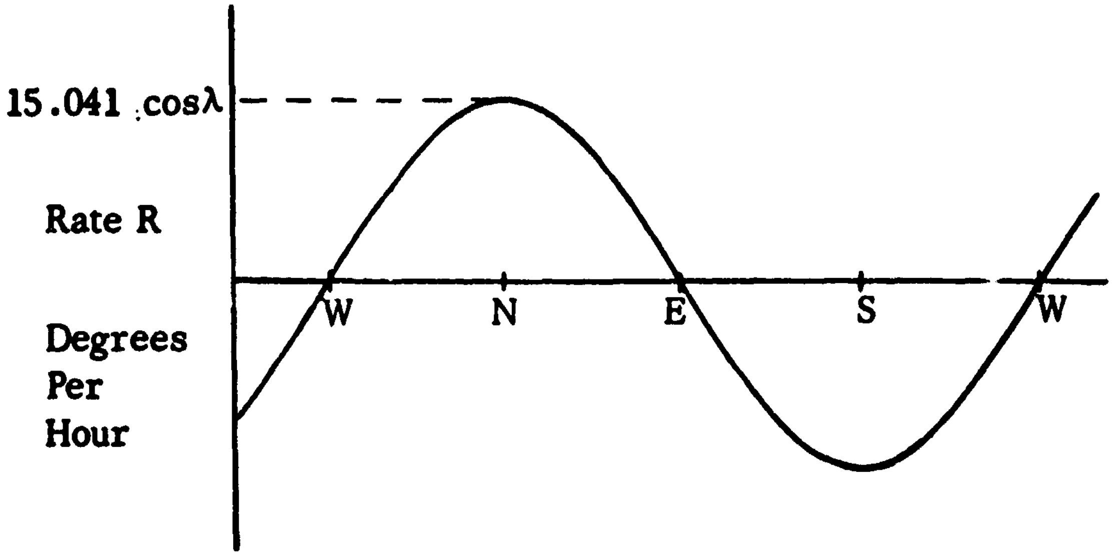


Fig.3(a) Horizontal component R of earth's rotational rate against azimuth

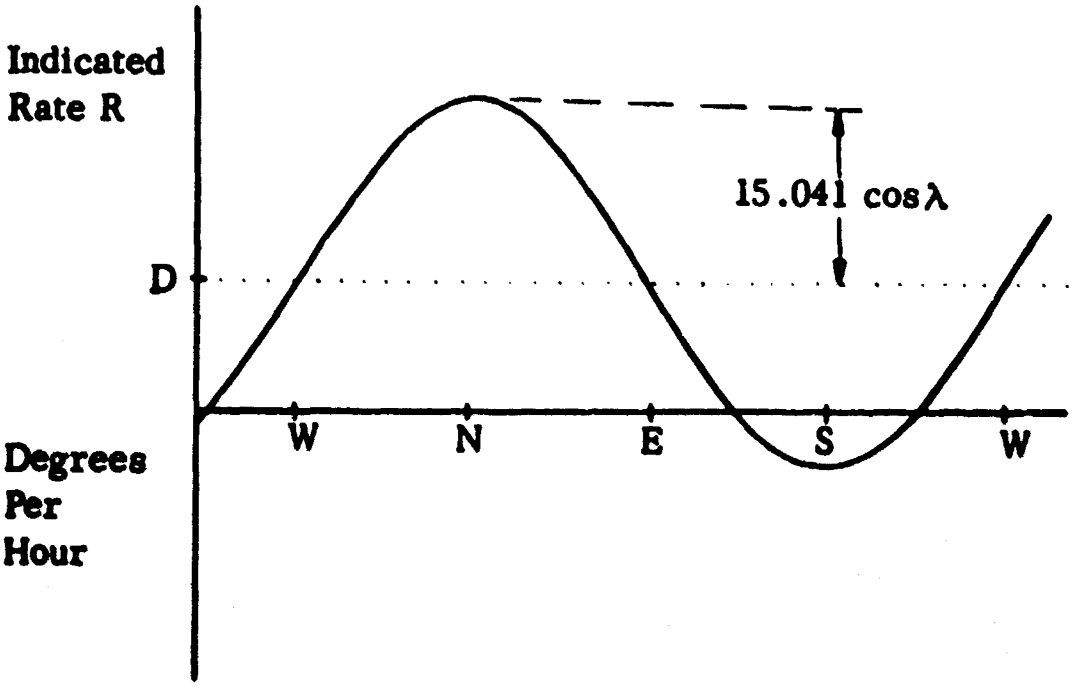


Fig.3(b) Rate R' indicated by a gyro with fixed error D

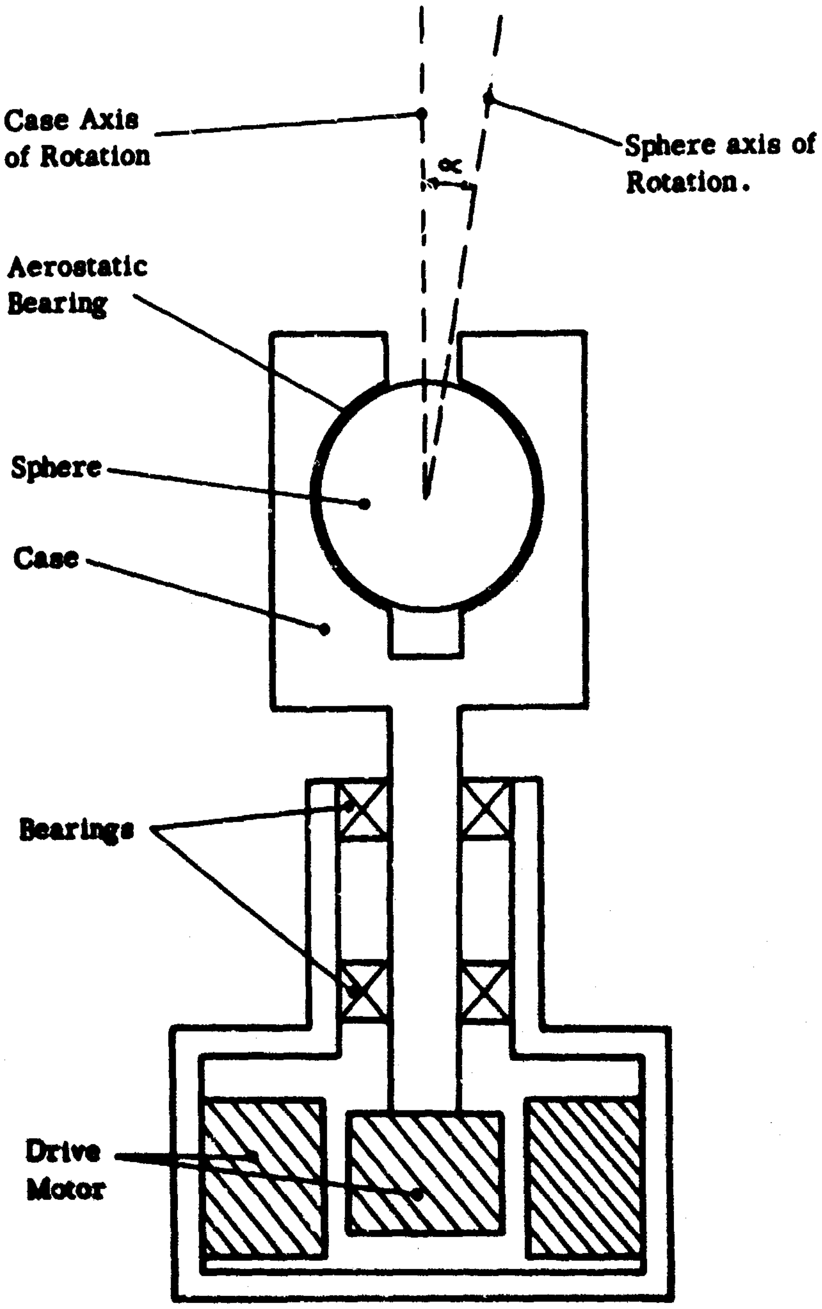


Fig.4 A two-axis rate gyro for North-finding. (Spherical bearing gas supply not shown)

PART II

SYSTEM PARAMETERS AS CONSIDERED IN THE DESIGN OF A SMALL INERTIAL NAVIGATION COMPUTER

by

J.Kydd

Litton Systems (Canada) Limited,
Rexdale, Ontario, Canada.

SYSTEM PARAMETERS AS CONSIDERED IN THE DESIGN OF A SMALL INERTIAL NAVIGATION COMPUTER

J. Kydd

1. INTRODUCTION

The system design of a small digital inertial navigation computer (DAN Mk1) will be discussed. The physical layout of the computer will be described briefly. Special attention will be paid to the factors and trade-offs which established the necessary word size, storage capacity, computing speed, and the interaction of sensor characteristics with those of the computer. The computer operates with whole binary numbers and with a fixed binary point. It has been described briefly elsewhere[1,2]. Figure 1 is a recent photograph. The computer exploits the CORDIC algorithms for trigonometric computation[3-6].

The following problems have been programmed:

(1) The inertial navigation problem with system axes chosen to be unipolar wander-angle.

Present position longitude and latitude (λ_p, ϕ_p) are computed from x and y accelerometer inputs.

(2) Course and distance to current destination or to alternate destination. The number of possible alternate destinations is four.

(3) The definition of flight-path. (Ground speed, ground track angle).

(4) Time-to-go and estimated time of arrival (ETA).

(5) Path control. (Distance cross-track, steering error signal).

The problems listed apply to "Navigate" mode. The mechanization is that of an inertial navigation wander-angle system and mainly follows the prescription of the Avionics Laboratory at Holloman Air Force Base[7]. There is also a program to align the system on the ground.

In navigation mode, the control loop containing the computer consists of the following: platform x, y, and z accelerometers - x, y, and z voltage to frequency converters (quantizers) - computer x and y velocity registers - computation of x, y, z angular space rates - computer Δ-modulation circuits - platform gyros. This loop will be traced through to establish the interaction of computer iteration-rate with input and output circuits. Special attention will be paid to scaling problems, particularly of double-precision operations used to calculate Coriolis terms and direction cosines. Figure 2 is a system diagram giving the flow of data (relevant to the above control loop) through the computer.

Referring to the figure, it is evident that the main system inputs are sign and magnitude signals of Q^X, Q^Y, Q^Z. These are, in their turn, outputs of voltage-to-frequency converters (also called "quantizers"). The basic circuit of a quantizer is as Figure 3. The d.c. inputs to the quantizers come from the corresponding accelerometers on the inertial platform. The other system input, h_B, is barometric altitude from the air-data system. It is used to compute vertical velocity, v_z, output from a baro-inertial vertical velocity computation loop. Another input to "input gating" is not shown in Figure 2. It is a data word read in on one line from a digi-switch in the control console.

For the moment, the physical form of the inputs and outputs will not be specified, but the information which they represent will be traced through the computer. Each input and output word has a definite address in the Scratch-Pad memory. Because the latter is a dynamic re-circulating delay line, the word-address is determined by the time at which a given word and its bits are available at the output of the line. The basic timing organization is represented in Figure 5.

The major-cycle is the electrical length of the delay-line loop. A typical word-address is defined by specifying the proper minor cycle, bit, and line, i.e. $m^5 G_1 B^2$ The word-address thus also specifies the input gating control suitable for entering a given word. All the input signals are synchronous with the major-cycle time-base of the computer. A given word consists of 26 bits and each bit of the given word occupies the specified bit-position in one to 26 groups. The total capacity of the memory (2 similarly organized lines) is 104 words of 26 bits. The 4 output lines which are permanently connected to the computer are shown on Figure 2. $\omega_x, \omega_y, \omega_z$ are the angular space rates of the platform used to torque platform gyros. ϵ_T is a steering error signal intended to keep own-craft on the true course to target. These variables are available to the output on one line each in the form of Δ-modulation (to be discussed presently).

The information progresses down the delay-line memory and normally emerges from the Sorter to recirculate into the 'A' input. This is the re-circulation loop. The Sorter also has a facility to erase a given word from the re-circulating flow under control of the program. The Memory has random-access capability. The main function of the Sorter is to extract a word from any memory slot specified by a fetch address and place the word in one of four data positions, y, x, θ, k. This is required by the arithmetic unit which works the CORDIC algorithms. The main CORDIC operations which have been mechanized are specified in Figure 6.

The Buffer is a small re-circulating memory of 4-word capacity (y, x, θ, k). It provides an interface between the memory and the arithmetic unit. The four words in the Buffer transfer to the Shuffler at a suitable time. This data is the input to a CORDIC operation as specified by the program.

The Shuffler separates the data stream into two streams containing the same data but in different order. For a given operation, the data re-circulates back to the Shuffler, for a total of 26 passes through the variable delay. The output y, x, θ, k have now been generated and appear in the Buffer. The result is then transferred back into any memory slot specified by the return address portion of the instruction.

Transfer operations, of which there are several, are treated as a special case of Multiply. A Compare operation has also been mechanized.

The outputs shown on Figure 2 are computed from programmed torquing algorithms. These will be described later.

2. PROCESSING OF ACCELERATION DATA INPUT TO THE PROGRAM

The design equations of the quantizer are

$$R_{(t)} = \frac{A}{A_m} f_{CP} \tag{1}$$

$$\left.\begin{aligned} k_1 &= \frac{2\, f_{CP}\, \Delta B}{K\, A_m} = \frac{1}{RC} \\ k_2 &= \frac{4\, f_{CP}\, \Delta B}{V_f} = \frac{1}{R_f C} \end{aligned}\right\} \tag{2}$$

$$N' = \frac{2\, A_m}{A} \tag{3}$$

$$p = \left\lfloor \frac{q}{N'} \right\rfloor \tag{4}$$

$$n = \left\lceil \frac{A_m}{A} \right\rceil , \tag{5}$$

where f_{CP} = the repetition rate of the clock pulse train driving the quantizer (as shown on Figure 3).

A_m = the quantizer design maximum (of input acceleration, A).

$R_{(t)}$ = pulse repetition rate of quantizer outputs (typical waveforms are shown in Figure 4).

k_1 = static gain measured from input of quantizer to input of level detectors.

k_2 = static gain measured from feedback input of integrator to input of level detectors.

$+\Delta B$ = positive trip level (volts). When d.c. level at input of positive level detector is $\geq \Delta B$, then the detector output is unambiguously switched to its 1 level.

V_f = magnitude of d.c. level switched to its output by either feedback driver switch when $+1$ is on its input.

K = volts per g scale factor of accelerometer output (as connected to quantizer).

n = the breakout number of the quantizer output = number of frame-times to the first output pulse. (It is the first integer greater than A_m/A).

N' = the period number of the quantizer output.

p = the number of output pulses in q frame-times after breakout. It is the first integer lower than q/N'. The frame-time for breakout is counted as $q = 1$.

If T is the frame-time of the quantizer output (it is the pulse-width shown in Figure 4), then $T = 1/2f_{CP}$. Other useful relations are

$$T_D = nT \tag{6}$$

$$R_q = \frac{p}{qT}, \quad q = 1, 2, \ldots . \tag{7}$$

where T_D = the breakout or dead-time of the quantizer.

R_q = the q^{th} estimate of the output PRF.

Equation (7) is applicable only after breakout.

In this system, the quantizer clock pulse train is the major-cycle waveform of the computer.

This choice makes $f_{CP} = 1/(676\ \mu s) = 1.4793$ kHz and also gives any quantizer output pulse ready access to a suitable memory slot, Q^X, Q^Y, Q^Z (without any loss of quantizer output pulses).

Each of the three quantizer output lines have the same scale:

Mode	*Scale*
ALIGN/CALIBRATE	64×15 pps $= 0.96449$ metre $\sec^{-2}$
NAVIGATE	15 pps $= 0.96449$ metre $\sec^{-2}$

Only the NAVIGATE mode will be considered here.

The least velocity increment delivered by a quantizer to the computer (represented by one quantizer output pulse) is

$$\Delta v = \frac{A_m}{f_{CP}} = \frac{0.96449}{15} \text{ metre sec}^{-1} . \qquad (8)$$

2.1 Adequacy of Design Parameters of the Quantizer-Computer Interface

Whether or not the design parameters of quantizers as quoted above are adequate from the system point of view, can be answered by the following rough argument.

Test the quantizer by applying a step-function of acceleration of value equal to the accelerometer threshold, $A = 5.0 \times 10^{-5}$g. The true velocity at the output is $v_T = At$. The error $\Delta v = v - v_T$ is plotted in Figure 7(a). The time-base is established by computing the breakout time and period of the output waveform. Equation (8) gives the velocity increment for each pulse, $\Delta v \simeq 0.212$ ft $\sec^{-1}$, and $A_m \simeq 9.75$g .

Thus

$$T_D = nT \simeq 65.8 \text{ sec.}$$

$$N'T = 3.9 \times 10^5 \times 338 \times 10^{-6} \simeq 132 \text{ sec.}$$

The amplitude of Δv varies roughly between ±0.1 ft $\sec^{-1}$.

The plot of velocity error shown exhibits an error source; it is not the propagated value of the error. It is likely that a velocity error of the form shown is less detrimental to the system than is an initial velocity (offset) error. The latter type of error gives rise to a system position error[10] of

$$\Delta x = \frac{\Delta \dot{x}_0}{\omega_S} \sin \omega_S t , \qquad (9)$$

where ω_S is the angular Schüler frequency. This equation applies to a tangent plane system but it should apply approximately for short-time missions. If $\Delta \dot{x}_0 = 0.1$ ft $\sec^{-1}$, then the maximum $\Delta x \simeq 80$ ft. This is negligible compared to a position error propagation for an inertial system (short-time mission) of typically 1.25 nautical miles per hour of flight[11].

Figure 7(b) is a plot of the velocity quantization error for $A = 2.0g$. The input pulses are collected during the "fast loop" iteration time of the computer, T_F (calculated in the next section). The d.c. component of the waveform shown is about 1 ft sec^{-1}. If this number is used in the same rough argument made above, the maximum $\Delta x \simeq 800$ ft. In practice, the position-error due to this source should turn out to be considerably less than this.

3. SYSTEM PARAMETERS OF THE COMPUTER

3.1 Word Size

The word size had to be determined not only by the presently contemplated use (inertial navigation) but also other possible applications of the computer. It was assumed that angle numbers required a 20 bit word-length. This is suitable for a Stellar Inertial Doppler system, where stellar bearing and elevation have 2^{19} bit accuracy. The resolution out of 360° is 1.2 $\widehat{\text{sec}}$. It is true that most shaft-angle to digital transducers in use in most systems have not the accuracy indicated above.

An error analysis of the CORDIC algorithms used showed that a 26 bit word-length was necessary to achieve the order of accuracy indicated.

3.2 Iteration-rate of Programs

The computer updates the angular space rates of the platform $(\omega_x, \omega_y, \omega_z)$ at the highest computing rate.

The iteration-rate required may be computed roughly as follows:

$$T_1 = \frac{\min(\Delta\theta_y)}{\max(\omega_y)} = \frac{2\ \widehat{\text{sec}}}{60\ \text{deg hr}^{-1}} = 0.03\dot{3}\ \text{sec}\ .$$

where 2 $\widehat{\text{sec}}$ = minimum angular increment of the torquing waveform output from the computer.

60 deg hr^{-1} = the maximum angular space-rate (of ω_x, ω_y, ω_z) in the wander-angle system, NAVIGATION mode.

Each operation executed by this computer requires two instruction words (each of 32 bits). An analysis of the present programs shows the following allocation of instructions:

Constants	41
Align, Gyrocompassing, Calibration	132
Navigation	278
	451

The present Read-Only Memory (where these instructions are stored) has a capacity of 512 32 bit words.

The NAVIGATION program effectively consists of the following "fast" and "slow" blocks, computed in the order given: F, S1, F, S2, F, S3, F, S4. The fast block contains 22 operations, whereas each slow block contains 20 operations. Each operation requires 676 μs to execute.

The torquing variables ω_x, ω_y, ω_z are computed in the fast block. The iteration time for the fast block is thus $(168/4) \times 676 \times 10^{-6} \simeq 0.028392$ sec.

The same iteration-time also applies to the computation of torquing variables in ALIGN modes.

The computation of azimuth angles in a wander-angle system is a thorny problem with some connection to computer iteration-rate requirements.

In Figure 8, various azimuth angles are defined relative to true North (N), platform x-axis (x_P), and aircraft longitudinal axis (x_A). These angles are platform azimuth angle (ψ), true heading (ϕ), and platform wander-angle (β). The block (P) denotes platform azimuth servo. The block (C) denotes the digital computer and (C′) is another computation block (possibly not included in the computer). The problem is whether or not C′ should be part of the digital computer or part of an external "Heading" servo. A complete discussion would entail a comparison of the economics and physical characteristics of a digital heading servo (driven by β) or the use of synchro/digital and digital/synchro converters. There is considerable activity in the development of the latter devices at the present time[9]. Our Company at Woodland Hills, California, is also active.

The present system assumes that an external heading servo is used. There are no azimuth slews under computer (and hence gyro) control. Eventually, a mixture of the two alternatives may be used. Azimuth slewing in ALIGN mode may be done under synchro control as (B) in Figure 8, but the computation of $\phi = \beta - \psi$ in NAV mode may be done by the computer. The computation must be done at a fairly high rate. $\dot{\phi} = \dot{\psi}$ and ψ can change rapidly in a fast horizontal turn. The iteration-time required, if one allows a lag of 0.5 degrees between computed and true heading, is as follows:

$$T_1 = \frac{\Delta\phi}{\dot{\phi}} = \frac{0.5 \text{ deg}}{20 \text{ deg sec}^{-1}} = \frac{1}{40} \text{ sec} .$$

In the azimuth control of Coarse Levelling typified by (B) of Figure 8, the external servo requires to know whether the angle β equals 360° or equals 90°. This is required because alignment of the platform can take place with x or y axis North as required by a calibration procedure.

The NAVIGATION program updates navigation variables like those listed in the Introduction at 8.8 times per second.

4. INTERACTION OF THE COMPUTER WITH QUANTIZED INPUTS

In this section it is necessary to do some scaling. The convention is followed that a physical variable is represented by a symbol with a bar. For example, in the case of earth-rate,

> $\bar{\Omega} = (SF\Omega)\,\Omega$, where $(SF\Omega)$ is the scale factor and the value of Ω is a number in a register of the computer (Max $|\Omega| = 1$).

Velocity components in directions given by platform axes are computed typically as follows:

$$\bar{v}_x = \int_0^t [\bar{A}_x - \overline{(\omega_y + \Omega_y)v_z} + \overline{(\omega_z + \Omega_z)v_y}]\, dt \qquad (10)$$

and there is a similar equation for $\bar{v}_y$.

The term involving the acceleration input is approximated using Equation (1):

$$\int_0^t \bar{A}_x\, dt \simeq \frac{A_m}{f_{CP}} \int_0^t R_{(t)}\, dt \simeq \frac{A_m}{f_{CP}} \sum_1 \frac{N(T_P)}{T_P} T_P , \qquad (11)$$

where $N_{(T_F)}$ is the number of quantizer pulses collected in the "fast block" iteration-time, T_F. The value of T_F was given in Section 3.

Quantizer input pulses are added into a register of the computer, Q_x, which is reset every iteration time. However, each pulse from the quantizer line is not added into the least significant bit position of the register (2^{-25}), but rather into the position (2^{-16}).

This is done to force the scale factor relation $SFQ_x = SFv$.

$$\bar{Q}_x = \frac{A_m}{f_{CP}} N_{(T_F)} \tag{12}$$

$$Q_x = \frac{A_m}{f_{CP}(SFv)} N_{(T_F)}$$

$$\text{Min}\,\Delta Q_x = \frac{A_m}{f_{CP}(SFv)} = 2^{-16}\,.$$

This equation determines the SFv.

The corresponding finite-difference equation in machine variables is

$$(v_x)_i = (v_x)_{i-1} + (Q_x)_i + T_F \frac{SFv}{a} \left\{ (\omega_z+\Omega_z)_{i-1} (v_y)_{i-1} - (\omega_y+\Omega_y)_{i-1} (v_z)_{i-1} \right\}. \tag{13}$$

The scale factor of all angular velocities is (SFv/a) where a is earth-radius. However, there is a problem (described in the next section) when one tries to add the last (Coriolis) term into the v_x register.

5. "DOUBLE PRECISION" OPERATIONS IN NAV MODE

In the DAN computer program it is necessary to maintain certain numbers to an accuracy greater than is possible with the present word-length. The two problem areas where "double precision" is required are the following:

(i) Computation of Coriolis terms to be added into a velocity register. (See Equation (13)).

(ii) Computation of direction cosines from the corresponding rates of change.

The computation of Coriolis terms is considered here.

A typical Coriolis term is

$$C_x = (\omega_z + \Omega_z)v_y - (\omega_y + \Omega_y)v_z\,.$$

Consider the case when $v_z \simeq 0$. The scale factor of the C_x register is $(SFv)^2/a \simeq 9.13$ ft sec^{-2}. The maximum of $(\omega_z + \Omega_z)v_y \simeq 1.405$ rad ft sec^{-2}. Therefore the maximum $(\omega_z + \Omega_z)v_y \simeq 2^{-3}$.

The precision to which $(\omega_z + \Omega_z)v_y$ must be expressed can be estimated thus:

$$\frac{\Delta(\omega_z + \Omega_z)v_y}{(\omega_z + \Omega_z)v_y} = \frac{\Delta v_y}{v_y} + \frac{\Delta(\omega_z + \Omega_z)}{(\omega_z + \Omega_z)}$$

Take
$$0.0002 \simeq \frac{\Delta(\omega_z)}{\omega_z} \leqslant \frac{\Delta(\omega_z + \Omega_z)}{(\omega_z + \Omega_z)}$$

$$\frac{\Delta v_y}{v_y} \simeq 10^{-4} .$$

Therefore
$$\frac{\Delta(\omega_z + \Omega_z)v_y}{(\omega_z + \Omega_z)v_y} = 3 \times 10^{-4} \simeq 2^{-12} .$$

Thus the number $(\omega_z + \Omega_z)v_y$ should extend in the C_x register from about 2^{-3} to 2^{-15}. According to Equation (13), this number is then multiplied by $T_p(SFv/a)$ to produce a term suitable to add into the v_x register. Call this term v_x'. From data already given, one calculates

$$\text{Max } v_x' \simeq 2^{-19} .$$

This term v_x' would extend in the v_x register approximately from 2^{-19} to 2^{-3}. However, the register length extends only to the 2^{-25} position. Thus significant places would ordinarily be rounded off.

In a typical situation of each of the problems listed above, a small term multiplied by a factor $T_p(SFv/a)$ is to be added into another register at each iteration of the "fast loop". However, this would entail dropping significant places of the small term. The following double-precision operation is used instead.

Consider two registers A and B. A is construed to be more significant and B less significant. Let the small term be C_x and the variable in A be v_x. The object is to add C_x into v_x as described above. One way to do this is as follows:

(1) Accumulate C_x in the register B.

(2) Test B for significance in the A register.

(3) When B is significant, transfer the significant part of B to the A register.

(4) Reset B to the insignificant part of B.

Note that the A register receives the accumulation of small terms from B which are significant in A and immediately when they become significant. Bits in B which are not transferred are not lost; this procedure is not difficult to program on our computer but the actual programming cannot be given here.

6. COMPUTER-GENERATED PLATFORM-TORQUING WAVEFORMS

Output torquing waveforms on three output lines, $\omega_x, \omega_y, \omega_z$, are generated by operations on three registers in the computer named T_x, T_y, T_z.

The same generating algorithms are used in all modes whether Course Levelling, Gyrocompassing, or Navigation.

The algorithms are

$$(T_x)_{i+1} - (T_x)_i + K_\varphi \text{ sign } (T_x)_i = (\omega_x)_{i+1}$$

$$(T_y)_{i+1} - (T_y)_i + K_\varphi \text{ sign } (T_y)_i = (\omega_y)_{i+1}$$

$$(T_z)_{i+1} - (T_z)_i + K_\varphi' \text{ sign } (T_z)_i = (\omega_z)_{i+1} .$$

The transition $i+1 \leftarrow i$ signifies an elasped time of T_F, the iteration-time of the "fast loop" in the program. The functions written on the right-hand side of the equations are forcing functions for the waveform generation. They are the current values of the appropriate angular space rates required to torque the platform and as computed by the program.

The waveforms on the output lines are formed merely by sampling the sign of the T_x, T_y, T_z registers as appropriate and as indicated in Figure 9.

For illustration, the torquing equations of Coarse Levelling (in ALIGN mode) will be outlined.

6.1 Torquing in Coarse-Levelling

Earth-rates are taken to be zero in coarse-levelling. In the program earth-rates are loaded with zero values and non-zero values are first calculated at the start of fine-levelling.

The iteration time for torquing is the same as that calculated previously and is 0.028392 sec.

6.1.1 x and y Torquing in Coarse-Levelling

(The equations are the same for X-axis North or Y-axis North).

T_x, T_y, T_z are registers used to establish torquing output waveforms on three output lines W_x, W_y, W_z. For x and y, the algorithms mechanized are

Let "$Q_y \neq 0$" be denoted by Q_y

"$Q_y = 0$" be denoted by $\bar{Q}_y$ and so on.

The state of the input registers is then specified by

Case 1: $Q_y \; Q_x$

Case 2: $Q_y \; \bar{Q}_x$

Case 3: $\bar{Q}_y \; Q_x$

Case 4: $\bar{Q}_y \; \bar{Q}_x$.

These four cases are covered by the program as follows:

Case 1:

$$(T_x)_{i+1} = (T_x)_i - C_2 \,\text{sign}\, (T_x)_i - C_1 \,\text{sign}\, (Q_y)_i; \qquad C_1 = C_2$$

$$(T_y)_{i+1} = (T_y)_i - C_2 \,\text{sign}\, (T_y)_i + C_1 \,\text{sign}\, (Q_x)_i .$$

Case 2:

$$(T_x)_{i+1} = (T_x)_i - C_2 \,\text{sign}\, (T_x)_i - C_1 \,\text{sign}\, (Q_y)_i$$

$$(T_y)_{i+1} = (T_y)_i - C_2 \,\text{sign}\, (T_y)_i .$$

Case 3:

$$(T_x)_{i+1} = (T_x)_i - C_2 \text{ sign } (T_x)_i$$

$$(T_y)_{i+1} = (T_y)_i - C_2 \text{ sign } (T_y)_i + C_1 \text{ sign } (Q_x)_i \ .$$

Case 4:

$$(T_x)_{i+1} = (T_x)_i - C_2 \text{ sign } (T_x)_i$$

$$(T_y)_{i+1} = (T_y)_i - C_2 \text{ sign } (T_y)_i$$

The interval $i \leftarrow i+1$, marked by index i is the time between successive octals (112). This is the iteration-time computed in 3.2 above.

6.1.2 *z-torquing in Coarse-Levelling (X-axis North)*

Case 1:

$$Q_x \neq 0 \ . \qquad (T_z)_{i+1} = C_1 \text{ sign } (Q_x)_i$$

Case 2:

$$Q_x = 0 \ . \qquad (T_z)_{i+1} = (T_z)_i - C_3 \text{ sign } (T_z)_i \ .$$

For Y-axis N, replace Q_x in *Case 1* by Q_y.

6.1 3 *Scaling Relations*

According to the list of constants, $C_1 = C_2 = 100\, C_3$.

$$SFT_x = SFC_2 = SFT_y = SFC_1 = 6 \frac{SFv}{a}$$

$$SFT_z = SFC_3 = 100 \frac{SFv}{a} \ .$$

$\frac{SFv}{a}$ is the same as in NAV mode.

$$\text{Max } \tilde{W}_x = \text{Max } \tilde{W}_y = 360^0 \text{ hr}^{-1}$$

$$\text{Max } \tilde{W}_z = 60^0 \text{ hr}^{-1} \ .$$

Angle Increment of (W_x, W_y) waveforms corresponding to one iteration-time interval = $(\text{Max } W)\,\Delta T = 360 \ \widehat{\text{sec}} \ \text{sec}^{-1} \times 28 \cdot 078 \cdot 10^{-3} \text{ sec}$

$$\simeq 10.1 \text{ sec} \ .$$

Angle Increment of (W_z) waveform corresponding to one iteration-time interval = $(\text{Max } W_z)\,\Delta T \simeq 1.7 \ \widehat{\text{sec}}$.

Estimate of time required to develop a quantiser input pulse rate of 1 pulse per iteration (compare min θ = 12.5 $\widehat{\text{min}}$ as calculated below) is

$$\frac{12.5 \times 60 \ \widehat{\text{sec}}}{360 \ \widehat{\text{sec}} \ \text{sec}^{-1}} \simeq 2 \text{ sec} \ .$$

6.1.4 *Computation of C_1*

Remember $C_1 = C_2$ and $(SFv/a) = 0.6606504 \times 10^{-3}\ sec^{-1}$.

Using $1^0 = 17.453292$ milliradian,

$$C_1 = C_2 = \frac{60^0\ hr^{-1}}{\frac{SFv}{a}} = \frac{60 \times 17.453292 \times 10^{-3}}{0.6606504 \times 3600 \times 10^{-3}}$$

$$= 0.440305795\ .$$

6.1.5 *Input Sensitivity*

There are two input pulse trains from two quantizers Q_x, Q_y. Each input pulse is added into a memory slot marked Q_x, Q_y as appropriate, but at the 2^{-16} point of the register instead of 2^{-25}.

In align and calibrate modes each quantizer pulse (input to the computer) has the following calibration in terms of platform accelerometer output, KA_x or KA_y :

$$64 \times 15 \text{ pulses sec}^{-1} = 0.96449 \text{ m sec}^{-2}\ .$$

6.1.6 *Interpretation of Input Sensitivity*

The following calculations are not required in the present program for scaling but they do establish useful interpretations.

Least velocity increment input to DAN corresponding to 1 input pulse from quantizer is

$$\frac{0.96449}{64 \times 15} \text{ m sec}^{-1}\ .$$

Scale factor of Q_x, Q_y registers

$$SFQ_x = SFQ_y = \frac{0.96449 \times 2^{16}}{64 \times 15} \text{ m sec}^{-1}\ ,$$

using $2^{16} = 65536$ exactly

$1 \text{ m} = 3.2808399 \text{ ft}$

$SFQ_x = SFQ_y = 216.0188 \text{ ft sec}^{-1}$.

Maximum number of pulses entering Q_x or $Q_y \doteq 42$.

Maximum of mean acceleration (averaged over an iteration interval) is

$$\left(\frac{\Delta v}{\Delta t}\right)_1 = \frac{216.02928 \times 42 \times 2^{-16}}{42 \times 676 \times 10^{-6}}$$

$$= 4.87625 \text{ ft sec}^{-2}\ .$$

This corresponds to an initial level misalignment given by

$$g \sin \theta \gtrsim 4.88 \text{ ft sec}^{-2}\ .$$

i.e. $\text{Max } \theta \gtrsim 8^\circ 48'$.

Minimum Q_x or Q_y (taken as 1 pulse every iteration):

$$\text{Min } (\overline{\Delta v})_1 = (SFQ_x)\, 2^{-16}$$

$$\text{Min}\left(\frac{\Delta v}{\Delta t}\right)_1 = \text{Min (acceleration)} = \frac{1}{42} \text{ of the value above.}$$

Hence Min $\theta \simeq 12.5\ \widehat{\text{min}}$, where this minimum corresponds to 1 pulse every iteration.

A particular case of coarse-levelling (when both quantizer registers Q^X, Q^Y are greater than zero) is illustrated in Figure 9. The output waveforms are single-sided. These are converted to double-sided current waveforms (with polarity as marked on the diagram) in the gyro-torquing amplifiers on the platform. The appropriate scaling levels are written beside the waveforms on the diagram. These levels are chosen by switching in appropriate resistors in the torquing amplifiers.

In coarse-levelling the control is bang-bang. The torquing is full-scale in magnitude with sign determined by the sign of the quantizer registers. These are reset every iteration as in NAV mode.

6.2 Δ-Modulation of Torquing Waveforms

The essential idea of Δ-modulation is that the information on the waveform is carried by the + and - elements as labeled in Figure 9. The time-averaging torquing rate, ω_x say, is related to this waveform, $i_{(t)}$, as follows:

$$\text{ave } \omega_x = \frac{K \int_0^{T_j} i_{(t)}\, dt}{T_j} ,$$

where T_j is the period of the waveform, and $i_{(t)}$ is the current through the gyro-torquing coil. A typical value of K is 27 ± 0.02 deg $hr^{-1}\ ma^{-1}$.

This equation can be rewritten as

$$\text{ave } \omega_x = \frac{KIT}{T} \frac{(n_j - k_j)}{N} ,$$

where T is the time interval of a ± element in the waveform, n is the number of + elements in N, k is the number of - elements in N, and N is the total number of elements considered.

If the maximum ω is represented by a completely "filled-in" waveform (no pulses), then $n = N$ and $k = 0$. Thus max $\omega = KI$.

The current I is thus determined by the maximum torquing level in degrees per hour.

The basic assumption made to justify the use of this type of Δ-modulation in gyro torquing is that the gyro will define a space-angle corresponding to the input space-rate.

A concise derivation of the torquing algorithms of Section 6 and their relation to Δ-modulation is given in the next section.

A more complex example of waveform generation is plotted in Figure 11. The driving function is $\bar{\omega}_x = -60 + \hat{\Omega}_{North}$ and the example is pertinent in Fine-Levelling. It is instructive to compute ω_x from the waveform and compare it with the driving function (Table I).

TABLE I

Computation of ω_x versus Time from Waveform of Figure 11

$$\omega_{x(N)} = 60\ \text{hr}^{-1}\ \frac{[n_{(N)} - k_{(N)}]}{N}$$

t		Cumulative Sum*	Ratio of $60^\circ\ \text{hr}^{-1}$	Degrees hr^{-1}	$-60 + \bar{\Omega}_N$ deg hr^{-1}
Number of iteration intervals	Sec				
6	0.17	-6	-1	-60	-49.128
7	0.199	-5	-5/7	-42.86	
17	0.483	-15	-15/17	-52.941	
18	0.511	-14	-14/18	-46.667	
29	0.823	-23	-23/29	-47.586	
40	1.13	-32	-32/40	-48.000	
51	1.45	-41	-41/51	-48.235	
62	1.76	-50	-50/62	-48.387	
73	2.07	-59	-59/73	-48.493	
139	3.94	-113	-113/139	-48.777	
151	4.28	-123	-123/151	-48.874	
437	12.4	-357	-357/437	-49.016	

* Count $\begin{matrix}1\\-1\end{matrix}$ for $\begin{matrix}+1\\0\end{matrix}$

The waveform of Figure 11 was actually plotted out to 437 intervals but the portion shown is sufficient to illustrate the principle. A plot of the percentage error versus time is shown in Figure 12 (data taken from Table I). This delay in establishing a precise angular velocity for the gyro input is not deleterious to fine levelling or gyrocompassing.

6.3 Adequacy of Design Parameters of Torquing-Computer Interface

This cannot be answered by simple arguments but requires simulation studies using suitable models of the computer plus platform loop. We have no results to offer at the present time.

However, the following simple argument leads one to believe that the time-intervals for torquing, as given here, are adequate.

Suppose that steady torques have been applied to the platform gyros and that a steady state has been reached as at the end of align and gyrocompassing. At some point in time $(t=0)$ a new angular velocity forcing function is applied in the computer.

In line with the assumptions already made,

$$\theta_{x_N} = \sum_{i=1}^{N} \omega_{x_i} T_F$$

$$\omega_{x_i} = \omega_{x(i-1)} + \dot{\omega}_x T_f$$

$$\theta^c_{x_N} = 60 T_F \sum_{i=1}^{N} \text{sign } (T_x)_{i-1}$$

$$(T_x)_i = (T_x)_{i-1} - 60 \text{ sign } (T_x)_{i-1} + \omega^c_{x_i} .$$

Take $\omega^c_{x_i} = \omega_{x_i}$, the superscript denoting values established by the computer.

Take $\dot{\omega}_x = 2.21\ \widehat{\sec}\ \sec^{-2}$ (based on maximum acceleration = 7g).

$$(T_x)_0 = -30 \text{ deg hr}^{-1}$$

$$60\ T_F = 1.71\ \widehat{\sec}$$

$$\dot{\omega}_x T_F = 6.28 \times 10^{-2}\ \widehat{\sec}\ \sec^{-1} .$$

θ_{xN} is the *change* in space angle due to the contribution of space rate ω_x after $i = 0$.

Consider first that ω_x is 30 deg hr^{-1}. ($\dot{\omega}_x = 0$, $i > 0$.) The results are plotted in Figure 10. For θ_x, the solid ramp is θ_{xN} and the histogram is θ^c_{xN}.

If we now consider the effect on the T_x pattern of Figure 10 made by including $\dot{\omega}_x$, the following formula emerges:

$$\delta(T_x)_i = \left(\sum_{j=1}^{i} j \right) \dot{\omega}_x T_F ,$$

which gives the shift in successive levels of T_x.

However, the shift in levels of T_x will not have any effect on θ^c_{xN} until the negative amplitude of T_x in Figure 10 has shifted up to the zero level. A value of N for which this will be true is given by

$$\frac{N(N+1)}{2} \dot{\omega}_x T_F = 30 \text{ deg hr}^{-1} .$$

With the values given, $N \simeq 30$ iteration intervals. The time-delay is $NT_F = 0.852$ sec. At that time, the change in θ (due to $\dot{\omega}_x T_F$) is only

$$\sum_{i=1}^{N} \dot{\omega}_x T_F^2 \simeq 0.054\ \widehat{\sec} .$$

The above argument established only that the torquing "mechanism" will likely follow the change from steady torquing rates to time-varying ones. The following derivation of the torquing algorithm is also instructive in this context.

Consider two types of finite difference, δ and ∇.

$$\nabla\theta_{xi} = \theta_{xi} - (\theta_x)_{i-1}$$

$$\delta\theta_{xi} = \theta_{xi} - \theta^c_{xi} ,$$

where θ^c_{xi} is effectively established by the computer on the output line in iteration-interval i.

Note that ∇ operates on iteration-index i, whereas δ operates at fixed i.

Consider the identity

$$\nabla \frac{\delta\theta_{xi}}{T_F} = \delta \frac{\nabla\theta_{xi}}{T_F} \tag{14}$$

and define

$$\delta\theta_{xi} = T_F \, T_{xi} . \tag{15}$$

Hence

$$\nabla T_{xi} = \delta\omega_{xi} = \omega_{xi} - \omega^c_{xi} .$$

ω^c_{xi} will be Δ-modulated so it will have the form $\omega^c_{xi} = 60 \text{ sign } (\;)$ when 60° hr^{-1} is the maximum torquing rate.

The sign function is determined as follows:

$$\text{If} \quad (\delta\theta_x)_{i-1} \geqslant 0 ,$$

$$\text{i.e.} \quad (\theta^c_x)_{i-1} \leqslant (\theta_x)_{i-1} ,$$

$$\text{we want } \omega^c_{xi} = +60 .$$

$$\text{If} \quad (\delta\theta_x)_{i-1} < 0$$

$$\text{i.e.} \quad (\theta^c_x)_{i-1} > (\theta_x)_{i-1}$$

$$\text{we want } \omega^c_{xi} = -60 .$$

Thus

$$(T_x)_i = (T_x)_{i-1} - 60 \text{ sign } (T_x)_{i-1} + \omega_{xi}$$

$$|T_{xi}| \leqslant 120 \text{ deg hr}^{-1} , \text{ all } i .$$

Thus Equation (15) gives

$$|\delta\theta_{xi}| \leqslant 3.4 \ \widehat{\text{sec}} .$$

(Fine-Levelling, Gyrocompassing and NAV mode).

Thus the difference between the two θ_x curves of Figure 10 should remain small, even when the instantaneous value of ω_x (i.e. ω_{xi}) changes from one iteration to the next.

Assume that the orientation of the platform has been determined to high precision by torquing with constant rates (as is done in fine align and gyrocompassing). Then it is reasonable to expect that this high precision will be maintained and that the system will follow the changing space rates established by navigation mode.

ACKNOWLEDGEMENT

The writer takes this opportunity to thank Mr L.A.Borth, Director of Engineering, Litton Systems (Canada) Limited, for permission to publish this paper.

REFERENCES

1. *Product Data Sheet, No. 120.* Litton Systems (Canada) Limited, 25 City View Drive, Rexdale, Ontario.

2. *Design Aspects of a Digital Airborne Navigation Computer.* Electronics and Communications (Canadian), February 1966, p.21, Errata, *ibid,* April 1966, p.38.

3. Volder, J.E. *The CORDIC Trigonometric Computing Technique.* Institute of Radio Engineers Transactions, EC, 330, September 1959.

4. Daggett, D.H. *Decimal-Binary Conversions in CORDIC.* Institute of Radio Engineers Transactions, EC, 335, September 1959.

5. Parini, J.A. *Divic Gives Answers to Complex Navigation Questions.* Electronics, 105, September 5, 1966.

6. Roth, W.F. et al. *Hyperbolic Coordinate Converter, Final Report.* Lear Siegler, Inc. US Department of Commerce, AD 619 299.

7. Detachment No.1, Avionics Laboratory, Air Force Systems Command, Holloman Air Force Base, New Mexico.

   (a) *Final Engineering Report, Unipolar Mechanization for Inertial Navigation.* TR MDC-TDR-64-2359, 28 July 1964.

   (b) *Equations for Mechanizing the Low-Cost Inertial Navigation System.* Technical Memorandum No.1, March 1964.

8. REFERENCE NOT AVAILABLE.

9. Hyatt, G.P. *Solid-State Synchro-to-Digital Converter.* AFIPS Conference Proceedings, Vol.31, 1967, pp.269-279.

10. Kachickas, George A. *Error Analysis for Cruise Systems.* In *Inertial Guidance,* edited by G.R.Pitman, 1962.

11. Holm, Robert J. *The 1965 Evaluation of a Litton LN12 Inertial Navigation System in Civil Air Transport Operations.* Navigation, Vol.14, No.1, 1967, p.105.

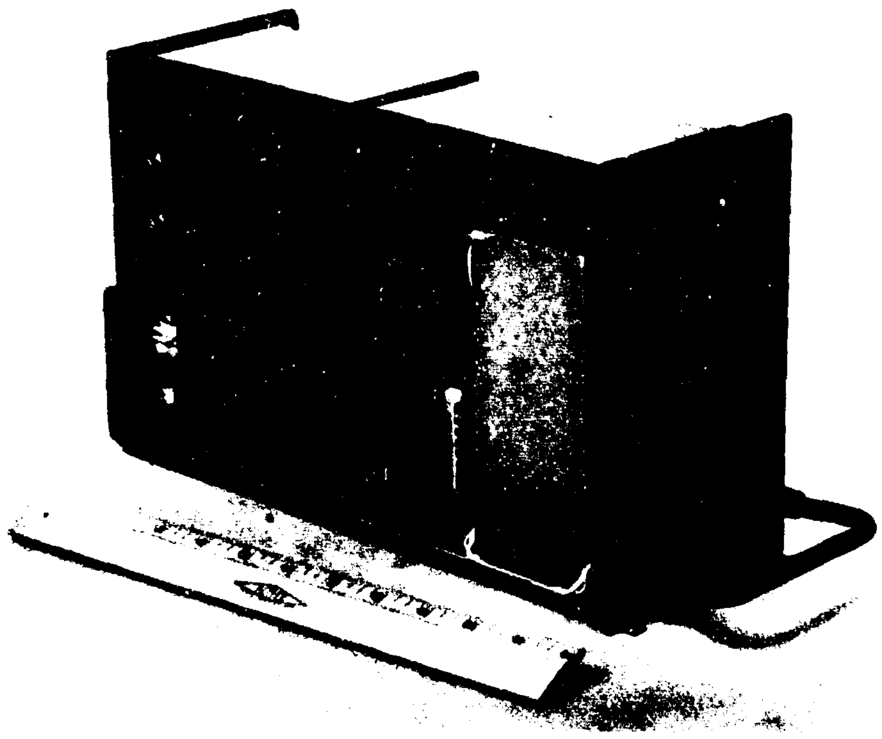


Fig. 1 DAN Mk1 computer

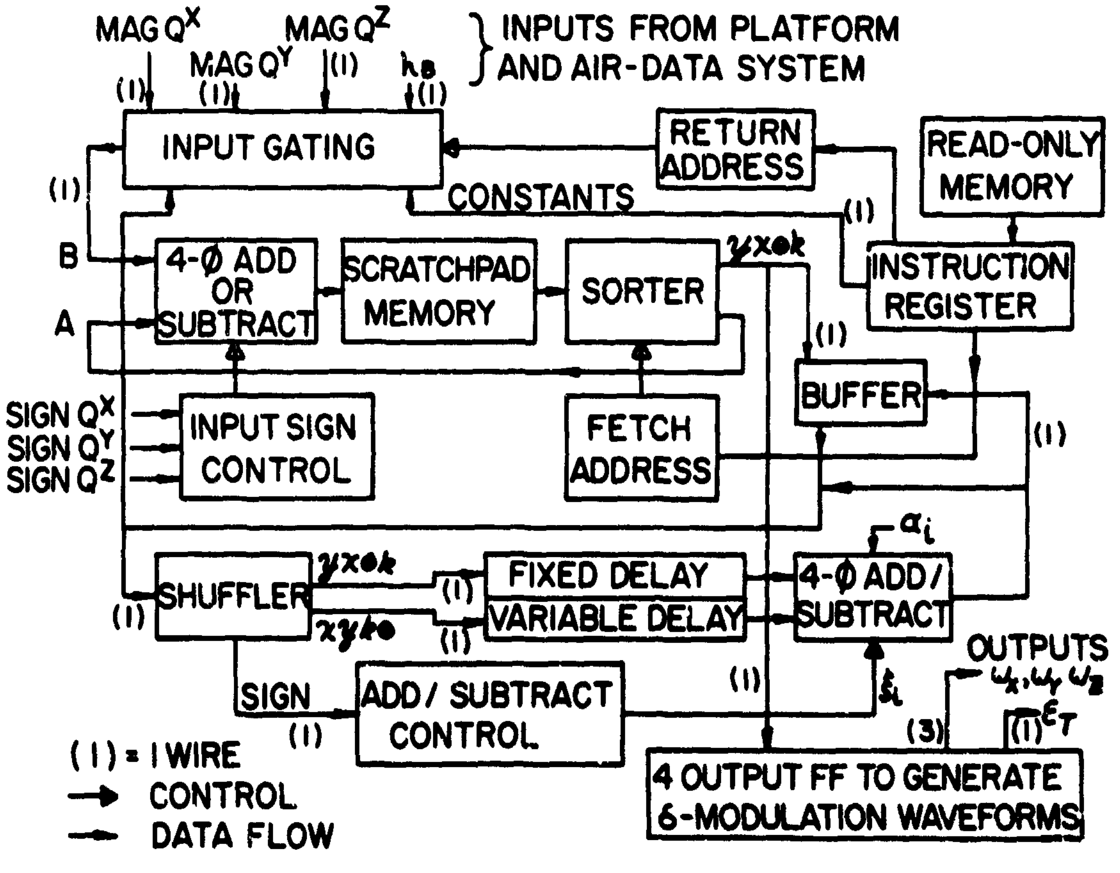


Fig. 2 System diagram showing relation of permanently-connected inputs and outputs to the internal data flow

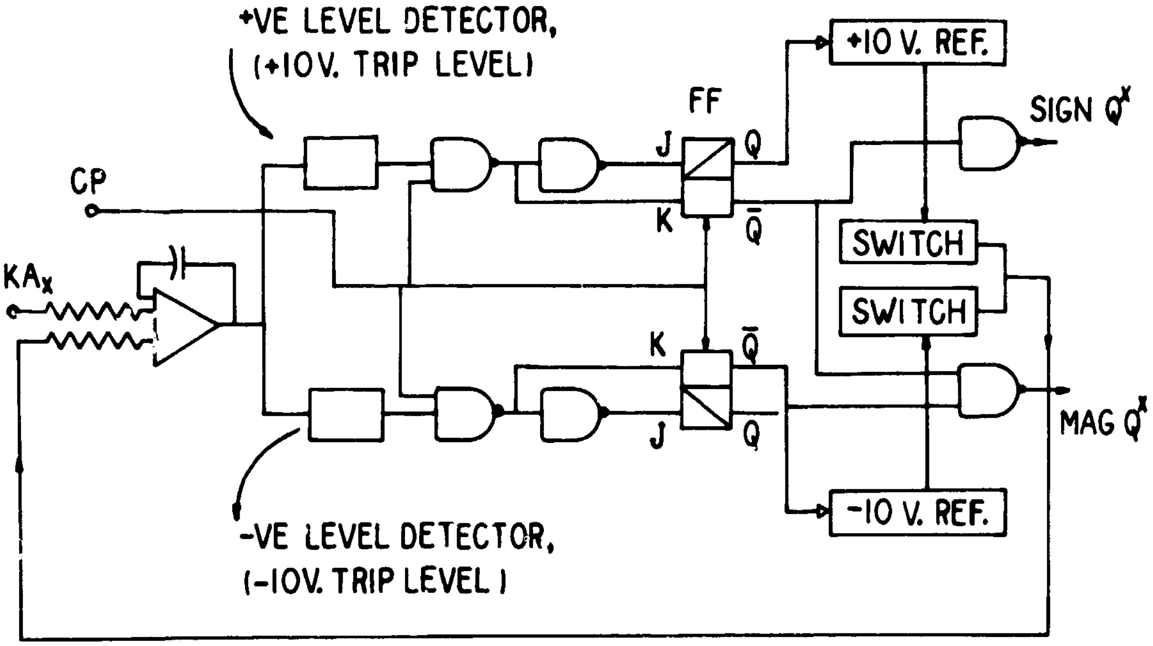


Fig.3 ‘D.C. input” to “pulse-rate output” converter

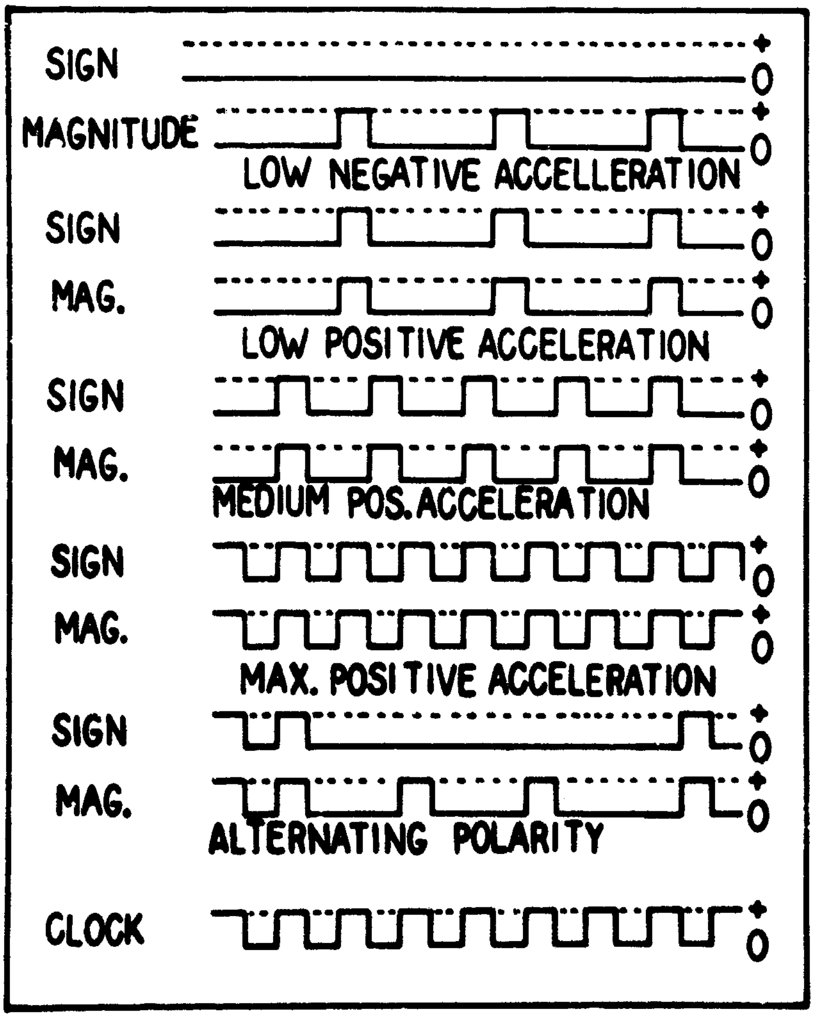


Fig.4 Typical output waveforms of voltage-to-frequency converter

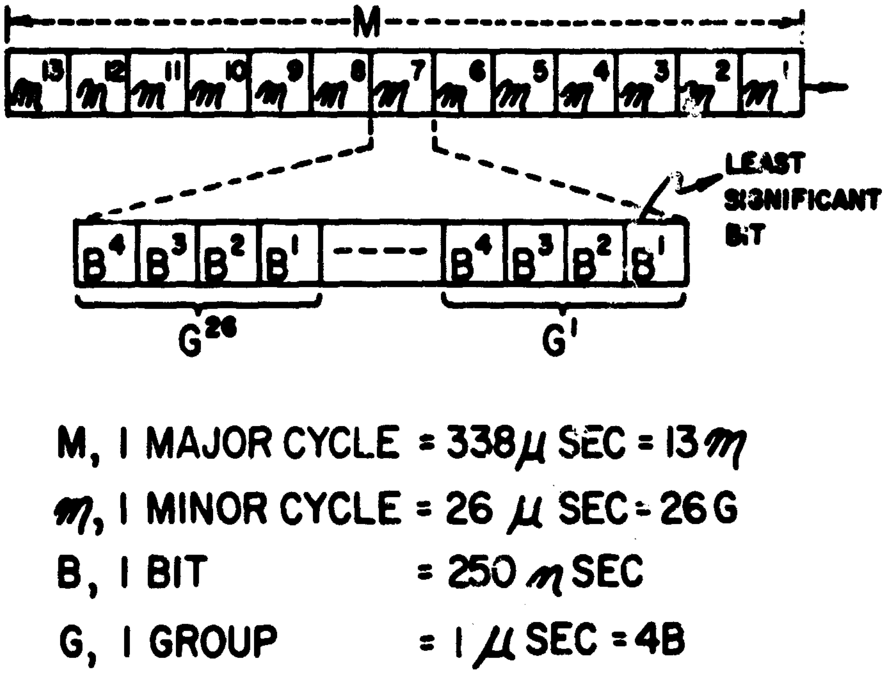
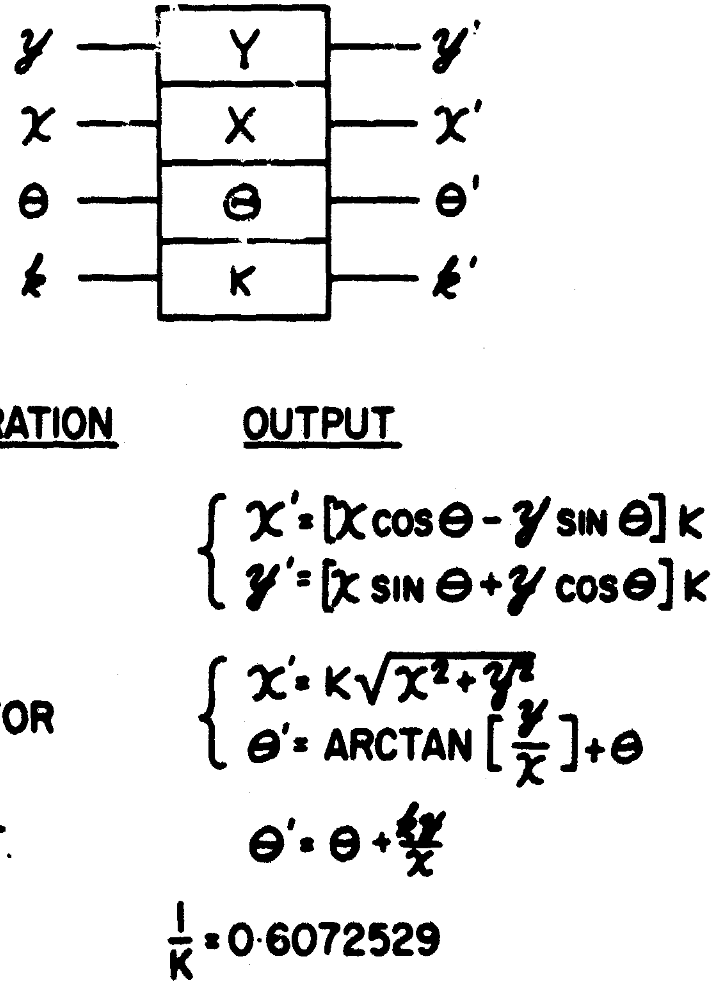

M
m^{13} m^{12} m^{11} m^{10} m^{9} m^{8} m^{7} m^{6} m^{5} m^{4} m^{3} m^{2} m^{1}
LEAST SIGNIFICANT BIT
B^4 B^3 B^2 B^1 ---- B^4 B^3 B^2 B^1
G^{26}
G^1
M, 1 MAJOR CYCLE = 338 μ SEC = 13 m
m, 1 MINOR CYCLE = 26 μ SEC = 26 G
B, 1 BIT = 250 n SEC
G, 1 GROUP = 1 μ SEC = 4B

Fig.5 Basic timing organization of digital computer DAN Mk1 based on re-circulating memory



OPERATION	OUTPUT
ROT	$x' = [x\cos\theta - y\sin\theta]K$ $y' = [x\sin\theta + y\cos\theta]K$
VECTOR	$x' = K\sqrt{x^2+y^2}$ $\theta' = \text{ARCTAN}\left[\frac{y}{x}\right] + \theta$
MULT.	$\theta' = \theta + \frac{ky}{x}$

$$\frac{1}{K} = 0{\cdot}6072529$$

Fig.6 CORDIC operations mechanized on digital airborne navigation computer

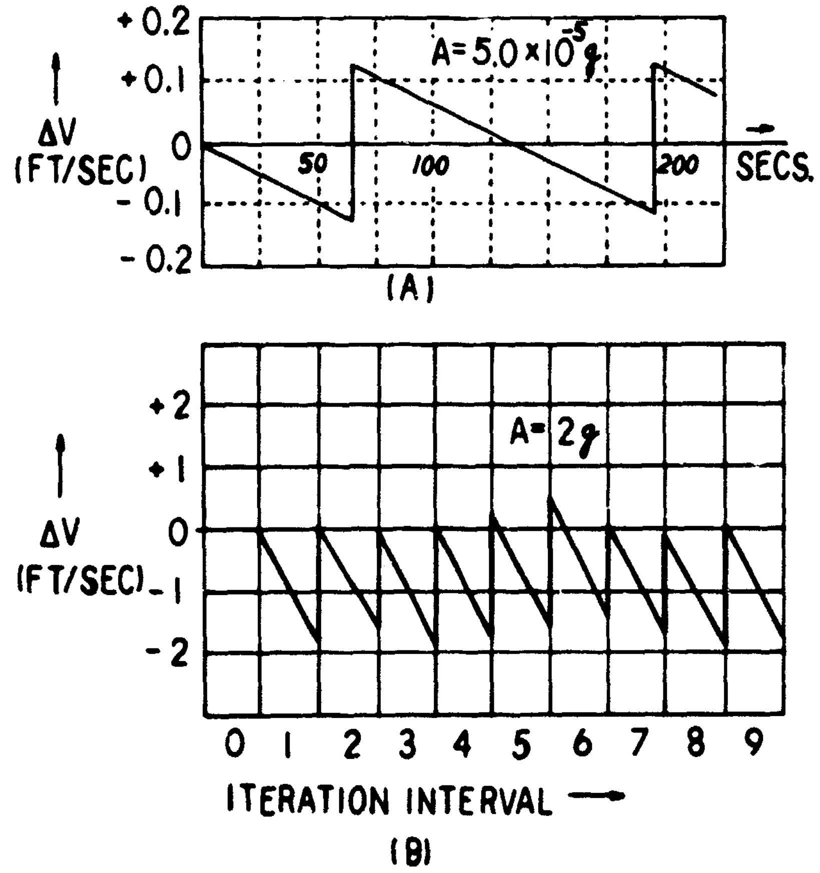


Fig.7 Quantizer velocity error due to quantization: $A = 5.0 \times 10^{-5}g$ and $A = 2g$

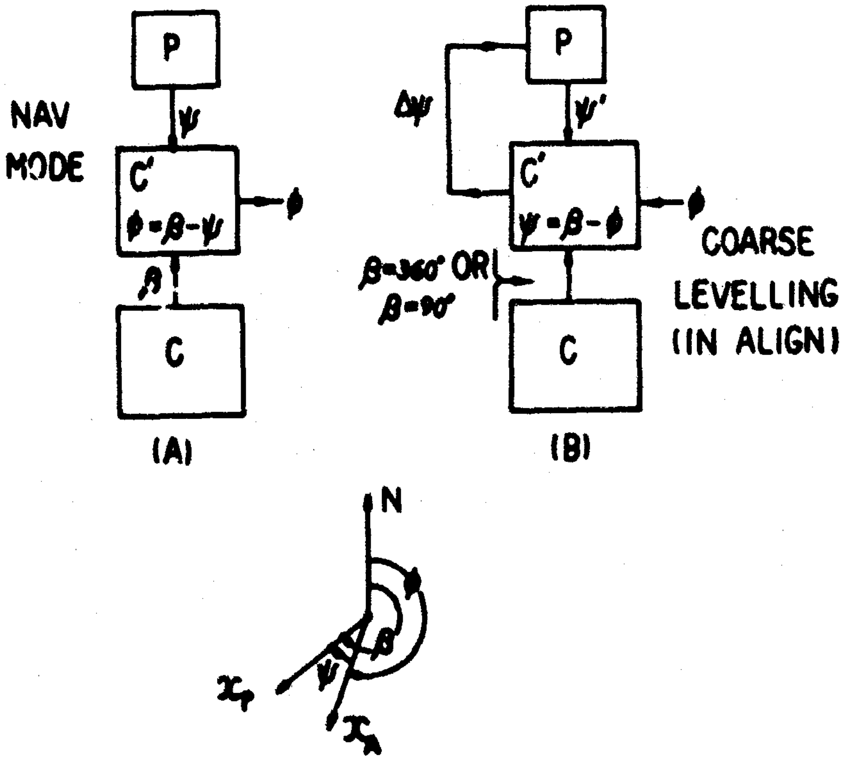


Fig.8 Computation of azimuth angles in a wander-angle system

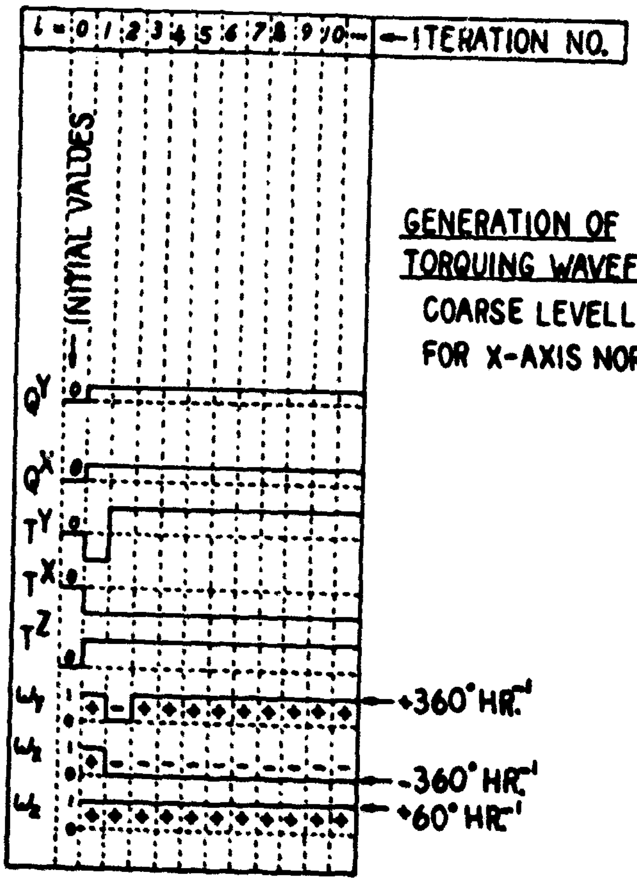


Fig.9 Generation of torquing waveforms: Coarse Levelling for X-axis North

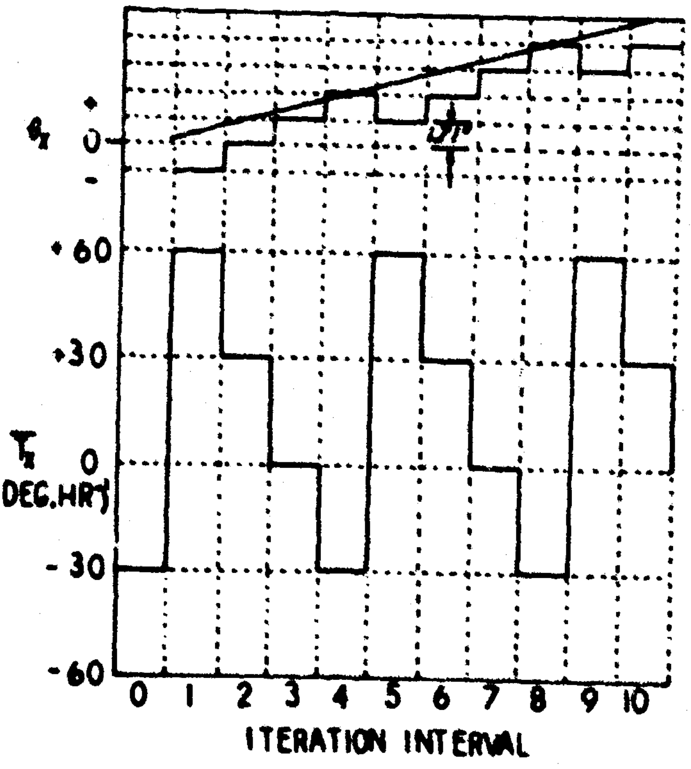


Fig.10 Time-delay of platform angle θ_x

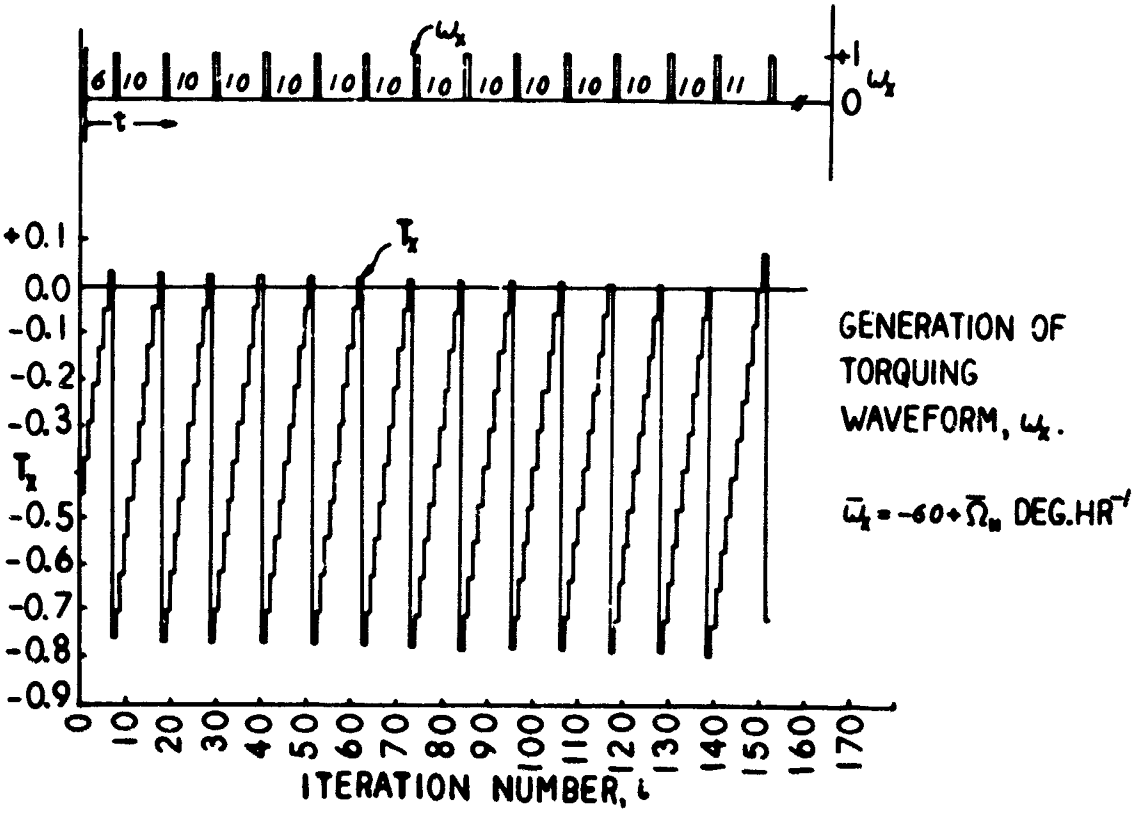


Fig. 11 Generation of torquing waveform ω_x

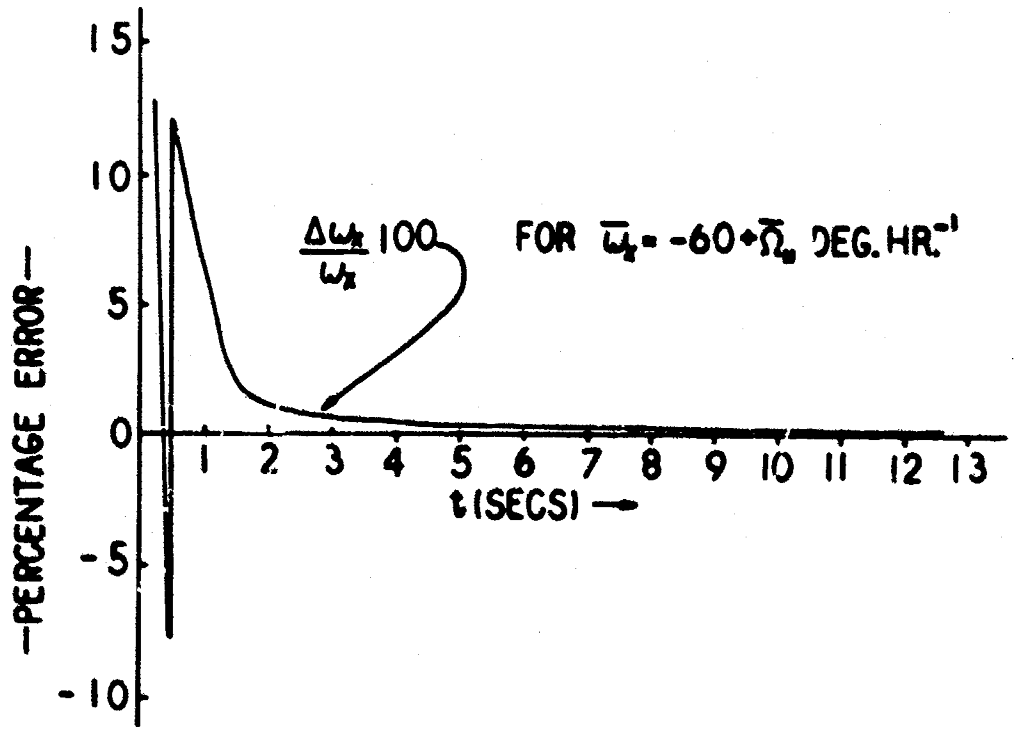


Fig. 12 Time required to develop accuracy of torquing waveform

THE APPLICATION OF STATISTICAL ESTIMATION TECHNIQUES TO INERTIAL COMPONENT CALIBRATION

by

George T. Schmidt

Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

SUMMARY

The calibration of inertial instruments in component, system, and vehicle level tests is shown to be amenable to modern statistical filtering techniques provided adequate models of the processes involved can be found together with some method of solving the immense computation problems. The general form of any calibration system using statistical filtering is given together with techniques for simplifying the computational problem. In one application gyro test table data is processed by a filter and the results compared with a Fourier series analysis. In a second application the design, implementation, and test experience is given for a unique calibration filter used to align and calibrate the Apollo Guidance Navigation and Control System inertial platform while it is in a launch vehicle that is subjected to wind induced sway.

THE APPLICATION OF STATISTICAL ESTIMATION TECHNIQUES TO INERTIAL COMPONENT CALIBRATION

George T. Schmidt

1. INTRODUCTION

Two recent developments are leading to changes in the way inertial instruments are calibrated in component, system, and vehicle level tests. One of these developments is the new statistical filtering, prediction, and smoothing techniques which provide the theory to account for all random noises, measurement errors, and disturbances affecting the component under test and its environment. The other complementary development is the availability of more powerful computers, both the large digital machine and the vehicle on-board computer, which provide the computational capability to take advantage of modern time-variable filter theory. There are, however, several practical problems encountered when one attempts to implement these techniques. A major one is in obtaining an adequate statistical model of the physical systems involved in the calibration process. Another problem is the immense amount of computation required. These two problems are interrelated in that the more accurately the system is statistically modeled, the more difficult the computational problem becomes.

The plan for this paper is to first show the form of any calibration system that uses statistical filtering together with the basic filter equations. Two applications with solutions will then be presented. The first is the use of a large digital computer to process gyro test table data and provide knowledge of gyro drift together with generating a model of gyro behavior. The filter is formulated and actual gyro test data is processed; the results are compared with a conventional Fourier series data reduction. This problem poses very little computational difficulty because of the availability of a large digital computer.

The second application is to the alignment and calibration of an inertial platform in a launch vehicle that is subjected to wind-induced sway while on the launch pad. This application requires an engineering solution to the real time computation problem with a limited on-board computer. We also impose the requirement that the calibration procedure be independent of any external references. Only the outputs of the components on the platform are to be used for calibration. The problem solution is discussed in detail and is, in fact, the method used in the Apollo Guidance Navigation and Control System.

The technique applied is to use only the two horizontal accelerometers to calibrate the gyros. The computation problem in implementing the optimum filter is solved by pre-computing the time optimal gains, neglecting minor gains, and approximating these gains by simple polynomials and exponentials in the guidance computer. It is shown that the procedure can be used to measure the South and vertical gyro drifts to one meru accuracy in 10 and 40 minutes, respectively. The platform can also be aligned in 10 minutes; this information is used to level the platform and then calibrate the vertical accelerometer. By a series of test positions all inertial components can be calibrated. The implementation problems are discussed in detail and the practical solutions are presented. Among them is the problem of accelerometer dead-zones near null inputs.

2. FORM OF A CALIBRATION SYSTEM USING STATISTICAL FILTERING

The basic configuration of an inertial component calibration system is shown in functional form in Figure 1. The question is, "What should the computer program be that converts the measurements to the desired outputs?" In other words, given a set of noisy physical measurements and some knowledge of the statistical characteristics of the components involved (e.g., accelerometers on a gyro-stabilized platform), find the best estimate of the desired output quantities.

The maximum amount of information provided by the measurements in a system with unpredictable errors is the probability density function $p(\vec{x})$. The vector $\vec{x}$ represents the state of the entire system involved, including the sensors, the vehicle (or mounting for the sensors), and the environment. With the knowledge of all past measurements and the *a priori* assumptions of the statistical characteristics of the entire system, the most that can be known about the state of a system is the probability that it is in each incremental volume of possible states. In most cases the entire probability density function is not desired. What is usually wanted is a set of values that is, in some respect, the "best" estimate of the state of the system. This estimate is symbolized by $\hat{\vec{x}}$, and it is usually chosen so as to minimize the average of some function of the error in the estimate $\vec{e} = \hat{\vec{x}} = \vec{x}$. The most obvious estimate, and one that minimizes the mean-squared error, is the expected or mean value which is given by

$$\hat{\vec{x}} = E(\vec{x}|\vec{m}) = \int_{-\infty}^{\infty} \vec{x} p(\vec{x}|\vec{m}_1,\dots,\vec{m}_n) d\vec{x} , \qquad (1)$$

where $p(\vec{x}|\vec{m}_1,\dots,\vec{m}_n)$ is the probability density function conditioned by all past measurements.

The solution of Equation (1) would require the computation of the entire probability density function of all the state variables as a function of time. The computation would include the changes in the density function due to the dynamics of the system and the changes due to the information gained by new measurements. The techniques for accomplishing this computation are not, to the author's knowledge, well developed and, furthermore, would be impractical for any computer in the foreseeable future. However, if we can assume that the errors in the estimate can be represented by a linear system excited by uncorrelated noise, then the whole process represented by Equation (1) can be stated concisely by equations developed by Kalman[7,8].

3. MATHEMATICAL DEVELOPMENT

It is assumed that the state of the entire system, including sensors, mounting and environment, can be described by differential equations of the form

$$\dot{\vec{x}}(t) = \vec{f}[\vec{x}(t), \vec{c}(t), \vec{n}(t), t] , \qquad (2)$$

where $\vec{c}$ are known control inputs and $\vec{n}$ are white noises. For the assumptions to be made here, it can be shown that the control does not affect the form of the optimum filter[9,10]. Thus the control variables $\vec{c}$ will not be shown explicitly in the following discussion. It is assumed that measurements are made at discrete times according to the relation

$$\vec{m}(t_m) = \vec{h}[\vec{x}(t_m), \vec{u}(t_m)] , \qquad (3)$$

where $\vec{u}(t_m)$ are errors in the measurements that are uncorrelated between measurements. (Time correlated errors in the measurements would have to be included in the state vector $\vec{x}$.) Assuming that the optimal estimates are close enough to the true values so that higher order terms may be neglected, the optimum measurement process is given by Kalman's optimum linear filter. The derivations of these equations are given in References 7, 10 and 11.

The basic equations are

$$\left.\begin{array}{l} \left.\begin{array}{rcl} \hat{\vec{x}} &=& \hat{\vec{x}}' + E'H^T(HE'H^T + U)^{-1}\ (\vec{m} - \vec{h}(\hat{\vec{x}}',t)) \\ E &=& E' - E'H^T(HE'H^T + U)^{-1}\ HE' \end{array}\right\} \begin{array}{l}\text{at a}\\ \text{measurement}\\ \text{time}\end{array} \\ \left.\begin{array}{rcl} \dot{\hat{\vec{x}}} &=& \vec{f}(\hat{\vec{x}},t) \\ \dot{E} &=& FE + EF^T + N \end{array}\right\} \begin{array}{l}\text{between}\\ \text{measurements ,}\end{array} \end{array}\right\} \qquad (4)$$

where the prime indicates conditions that exist just before the measurement. The covariance matrix of errors E is defined by

$$E = \langle \delta\vec{x}\delta\vec{x}^T \rangle$$

where "$\langle\ \rangle$" represents the expected or mean value, and

$$F = \frac{\partial\vec{f}(\hat{\vec{x}},t)}{\partial\hat{\vec{x}}}, \qquad H = \frac{\partial\vec{h}(\hat{\vec{x}},t)}{\partial\hat{\vec{x}}},$$

$$U = \frac{\partial\vec{h}}{\partial\vec{u}} R \frac{\partial\vec{h}^T}{\partial\vec{u}}, \qquad N = \frac{\partial\vec{f}}{\partial\vec{n}} Q \frac{\partial\vec{f}^T}{\partial\vec{n}} .$$

The matrices R and Q are defined by

$$\langle\vec{u}(t)\vec{u}(\tau)^T\rangle = R\delta(t-\tau)$$

$$\langle\vec{n}(t)\vec{n}(\tau)^T\rangle = Q\delta(t-\tau) .$$

The optimum filter for a linearized system in which the higher order terms can be neglected is shown in diagram form in Figure 2.

4. GYRO UNIT TESTING USING STATISTICAL FILTERING

4.1 Test Configuration

The test we will consider is the so-called "IA vertical" test in which the gyro with its input axis vertical is mounted on a rotary table. The gyro output axis is West and the spin reference axis South. The gyro output signal is used to drive the rotary table by means of amplifiers and motors. The table rotates about the vertical according to the differential equation

$$da/dt = d + Whb\cos(a) + Whc\sin(a) + Wv , \qquad (5)$$

where a = table angle measured clockwise from West

d = gyro drift

Wv = vertical component of earth rate

Wh = horizontal component of earth rate

b = misalignment about the output axis

c = misalignment about the spin axis

and the signs of b and c are by the user's convention. Other types of servo tests are possible and are discussed in Reference 18.

In a conventional servo table test, the time it takes for the table to move each one degree increment is recorded for two revolutions of the table; this requires about 72 hours of testing. The usual data reduction technique is to fit a Fourier series to the data and identify the constant in the series as the gyro drift and the coefficients of the cosine and sine terms as the misalignment angles. Typically the autocorrelation function is also calculated for the data and one attempts to find an autocorrelation function for some process that fits the data; e.g. a ramp or perhaps a random walk[19]. The curve fits are usually done by a least-squares method. The purpose here is to give an alternative approach using recursive statistical filtering, compare the results from both methods and the advantages in each, and suggest some areas for further research. A functional view of the optimal calibration system is shown in Figure 3.

4.2 The Filter

The time it takes the table to go each one degree is recorded on tape. The tape is fed into a digital computer which has programmed into it the following filter.

Between each measurement the state vector is integrated according to:

$$\frac{d\hat{\vec{x}}}{dt} = \frac{d}{dt}\begin{bmatrix} a \\ d \\ b \\ c \\ e \end{bmatrix} = \begin{bmatrix} 0 & 1 & Wh\cos(a) & Wh\sin(a) & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} a \\ d \\ b \\ c \\ e \end{bmatrix} + \begin{bmatrix} Wv \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \qquad (6)$$

where e has been introduced to account for any bias in the measurement of table angle. The initial condition for the state vector integration is the estimated value after the last measurement incorporation.

The extrapolation of the covariance matrix between measurements is a bit harder to derive. Linearizing the differential equation for a we have

$$\delta\dot{\hat{a}} = \delta\hat{d} + Wh\cos(\hat{a})\,\delta\hat{b} + Wh\sin(\hat{a})\,\delta\hat{c} - Wh\hat{b}\sin(\hat{a})\,\delta\hat{a} + Wh\hat{c}\cos(\hat{a})\,\delta\hat{a}\ . \qquad (7)$$

The measurements of table angle, denoted by m, are assumed to be highly accurate; typically, they are true to within a few arc-seconds. Since the misalignment angles are small, one may neglect the last two terms in Equation (7) and the F matrix in the covariance matrix extrapolation,

$$dE/dt = FE + EF^T\ , \qquad (8)$$

is given by the matrix in Equation (6). The initial condition for the covariance matrix is the value after the last measurement incorporation. Equations (6) and (8) are integrated on the digital computer until the time of the next measurement (m). At that time the equations are updated according to

$$\vec{k} = E'\vec{h}(\vec{h}^T E'\vec{h} + ra)^{-1}$$

$$\hat{\vec{x}} = \hat{\vec{x}}' + \vec{k}(m - \vec{h}^T\hat{\vec{x}}')$$

$$E = (I - \vec{k}\vec{h}^T)E'\ ,$$

where ra is the variance of the measurement noise and $\vec{h}^T = (1,0,0,0,1)$. The integration process (Equations (6) and (8)) is then re-initiated.

4.3 Test Results

An "IA vertical" test was run on a gas-bearing gyro and the data recorded as described. A digital computer was programmed to filter the information. One full revolution of data in one-degree increments was processed. The initial condition for the state vector was zero and all initial cross-correlation terms were set zero. The initial diagonal of the covariance matrix was: 9×10^{-8} rad^2, 100 meru^2, 9×10^{-8} rad^2, 1×10^{-6} rad^2, and 1×10^{-10} rad^2. (A meru is approximately 0.015 degrees/hour.) The r.m.s. measurement noise was assumed to 1×10^{-5} rad.

The results of this test are plotted in Figures 4, 5 and 6. In Figure 4 we notice, first, that the estimated drift is, in two samples, very close to its steady state value of 12.158 meru. The estimated r.m.s. error $\surd(E_{22})$ is also plotted. In two samples it is down from its initial value of 10 meru to 0.6 meru; in 360 samples it is down to 0.003 meru.

In Figure 5 the b estimate is plotted, together with its estimated r.m.s. uncertainty. Apparently it takes about half a revolution before we have enough data to filter this quantity. The final $\surd(E_{33})$ is 0.001 mrad as opposed to its initial value of 0.3 mrad. Similar comments hold for Figure 6. The final value of $\surd(E_{44})$ is 0.0007 mrad as opposed to its initial value of 1 mrad. The final angle estimates in each case are 0.489 and 0.743 mrad, respectively. Finally, the estimate of e showed an insignificant amount of bias.

4.4 Comparison with Conventional Method

The same data presented in the example were fitted with a Fourier series. The drift and misalignment angles (12.17 meru, 0.544 mrad, 0.711 mrad) compared almost perfectly with those as determined by the filter (12.158 meru, 0.489 mrad, 0.743 mrad). One might ask the question, "Why bother with this sophisticated method if a Fourier series works?"

One possible advantage is that the data can be processed as it is received. One might imagine a number of test tables tied into a central computer which would calibrate a whole number of gyros at once and provide a best estimate of each gyro's current behavior. Another advantage is that the filter generates an r.m.s. uncertainty estimate in the estimated variables. If one has confidence in the statistics that must be assumed *a priori*, then in an *ensemble* sense, we can tell how good our test is. Furthermore, another advantage is in the ease with which additional state variables can be included in this formulation. We might like to try to add in a model for drift other than a bias; perhaps a ramp, an exponentially correlated process, or a random walk. These models are easily added to the filter.

One possible disadvantage may be found if we test a gyro that has sudden jumps to different steady-state levels. In all likelihood, the filter will probably take a long time to catch up to the actual gyro or it may never get there. This is the usual argument for a Fourier series expansion of any discontinuous function. In any case, the purpose here is to show the form of the system that could be used to reduce gyro test table data, not that any great improvement is guaranteed. What is needed is more research into this type of formulation.

5. METHODS FOR REDUCING COMPUTATION

Throughout the previous example it has been tacitly assumed that a large digital computer is available to do the computation. For this simple example five state variables were required. Since the number of state variables rapidly increases with the complexity of the problem, engineering solutions will usually be required, particularly if the digital computer has a limited capability. This is the case in an on-board vehicle computer. Generally there are two possible ways to reduce the computational burden.

The first method is to partition the total filter into smaller and simpler filters by neglecting the cross-correlations between dynamically unrelated variables. One might also neglect to implement minor gains for particular state variables; i.e. each state variable is updated by all measurements, even if the particular measurement has an insignificant effect on that particular state variable.

The second method of simplifying the optimum filter is through the use of pre-computed gains. If the system is linear, the gains at each measurement are only functions of time and the *a priori* assumptions of the statistics of the noises and the initial state. By specifying the measurement schedule or rate, the gains may be pre-computed and stored in the vehicle's computer. For a small number of measurements this is a practical solution to the problem of implementing an optimum filter. For a large number of measurements, the pre-computed gains are usually smoothly varying with respect to time and may be approximated by suitable functions that give an almost identical filter response as the true gains. The weighting function box in Figure 3 now contains simple functions of time.

These simplification techniques will be applied to the pre-launch calibration and alignment of an inertial platform in a spacecraft on top of a swaying launch vehicle and they are, in fact, the techniques used in the Apollo Guidance Navigation and Control System. Practical hardware and software problems that were involved will also be discussed in detail.

6. PRE-LAUNCH CALIBRATION AND ALIGNMENT

6.1 General

The inertial system to be calibrated and aligned includes gyroscopes and accelerometers. The known gravity acceleration is used to calibrate the accelerometers; the known vector rotation of gravity (earth rate) is used to calibrate the gyros. The exact quantities to be measured will not be considered at this point in developing the general optimum method of platform alignment to a local vertical coordinate system and measurement of the South and vertical gyro drifts. In this procedure, the platform is approximately aligned to the local vertical coordinates, then it goes inertial. The two horizontal accelerometer outputs (South and East) are used by the optimum filter to generate estimates of the relevant quantities by comparing the measurement of the rotation of the gravity vector with the known rotation rate.

The vertical gyro drift is the most difficult quantity to measure, since it causes only a third-order effect on the measured acceleration. Gyro failures can be closely associated with changes in drift due to acceleration of gravity along the input axis, so the pre-launch calibration of a gyro in a vertical position is highly desirable.

Since the estimates of the alignment and drift variables will depend on the measurement by the accelerometers of the rotation of the gravity vector in the inertial coordinates instrumented by the gyros, the major disturbances are the accelerometer quantization and the wind-induced sway of the launch vehicle. The model of the system for the optimum filter must include variables due to this sway. The complete filter must be simulated on a digital computer; it will be linear, so that by specifying the measurement schedule the optimum gains may be pre-computed. The gains will be approximated by functions that will be stored in the flight computer. The method for using this simplified filter in other platform positions, an illustration of a system test program, and practical hardware problems will be presented.

6.2 Models

The launch vehicle bending dynamics in the North-South and East-West directions are approximated by identical second-order systems. The wind causing the vehicle sway is assumed to be exponentially correlated, with a correlation time of $1/\lambda$ sec . The correlation function of the white noise required to produce a mean-squared value of missile sway can be found to be[13]

$$nw = \langle n(t)n(\tau)\rangle = \frac{\langle p^2\rangle 4\lambda\zeta\omega_n^3(\omega_n^2 + 2\lambda\zeta\omega_n + \lambda^2)}{\lambda + 2\zeta\omega_n} ,$$

where $\langle p^2\rangle$ is the expected mean-squared missile sway, $n(t)$ is the white noise generating the exponentially correlated wind, and ω_n and ζ are the natural frequency and damping ratio of the second-order approximation to the bending dynamics. The state vector for the sway variables in the South direction is

$$\begin{bmatrix} \dot{p}_s \\ \dot{v}_s \\ \dot{a}_s \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ -\lambda\omega_n^2 & -\omega_n^2-2\zeta\lambda\omega_n & -2\zeta\omega_n-\lambda \end{bmatrix} \begin{bmatrix} p_s \\ v_s \\ a_s \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ n(t) \end{bmatrix} \tag{9}$$

where p_s, v_s, a_s are the horizontal displacement, velocity, and acceleration in the North-South direction; the model for the sway variables in the East direction (p_e, v_e, a_e) has the same form. For the computer simulations in this paper, the preceding variables have the values $\langle p^2\rangle = 100\ \text{cm}^2$ (East and South), $\lambda = 0.1\ \text{sec}^{-1}$, $\omega_n = 2.09$ rad/sec, and $\zeta = 0.1$.

The orientation of the platform with respect to a local vertical coordinate system (vertical, South and East) is described by three angles (α, β, γ). If the platform axes (x, y, z) were rotated by $-\alpha$, $-\beta$, and $-\gamma$, the axes would coincide with the reference coordinates. The state vector equation for this substate is given by

$$\begin{bmatrix} \dot{\alpha} \\ \dot{\beta} \\ \dot{\gamma} \\ \dot{d}_x \\ \dot{d}_y \end{bmatrix} = \begin{bmatrix} 0 & 0 & \Omega_h & 1 & 0 \\ 0 & 0 & \Omega_v & 0 & 1 \\ -\Omega_h & -\Omega_v & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} \alpha \\ \beta \\ \gamma \\ d_x \\ d_y \end{bmatrix} + \begin{bmatrix} t_x - \Omega_v \\ t_y + \Omega_h \\ t_z + d_z \\ 0 \\ 0 \end{bmatrix} \tag{10}$$

which has been assumed to be, for angle magnitudes of interest, a valid representation of the general nonlinear platform dynamics[11]. Ω_h and Ω_v are the horizontal and vertical components of earth rate at the test site; d_x, d_y, d_z are the constant drifts for the vertical, South and East gyros; and t_x, t_y, t_z are the torquing rates (if any) applied to the gyros. It is assumed that the torquing rates, the components of earth rate, and the East gyro drift are known perfectly, so that the vector on the right represents known forcing $\vec{c}(t)$ and is independent of the state of the system.

Using the small angle approximations, the South and East accelerometer pulse rate outputs due to platform orientation in the gravity field may be written as

$$\begin{bmatrix} \dot{po}_s \\ \dot{po}_e \end{bmatrix} = g \begin{bmatrix} -\gamma \\ \beta \end{bmatrix} \tag{11}$$

where po_s and po_e represent the total pulse counts at some instant of time and g is the local gravity (cm/sec^2). A 1-cm/sec^2 pulse accelerometer quantization has been assumed.

Accelerometer pulse rates cannot be instantaneously measured but the total pulses can be counted which make up the output due to sway velocity and orientation in the gravity field. Inherent in these measurements, then, are quantization errors. The measurements are

$$\vec{m} = \begin{bmatrix} m_s \\ m_e \end{bmatrix} = \begin{bmatrix} v_s \\ v_e \end{bmatrix} + \begin{bmatrix} po_s \\ po_e \end{bmatrix} + \begin{bmatrix} n_m \\ n_m \end{bmatrix} . \tag{12}$$

The term "n_m" represents the quantization error at every sampling of the accelerometer pulse count registers. Although the measurement noise is not normally distributed with zero mean (a requirement for the optimum filter), and is, in fact, uniformly distributed, the quantization operation is assumed a normally distributed error. This viewpoint does not hinder the optimum filter's estimates of the alignment and gyro drift variables[10,11]. It does degrade the accuracy of the sway variables estimates, which is of little concern. Also

$$\langle n_m(t) n_m(t + \tau) \rangle = ra\, \delta(\tau) \tag{13}$$

The measurements are assumed to be made every second.

The state vector is 13-dimensional; the state vector differential equation is

$$d\vec{x}/dt = F\vec{x} + \vec{c} + \vec{n}(t) . \tag{14}$$

(See Figure 7 for $\vec{x}$, F , $\vec{c}$ and $\vec{n}$) .

With the derivation of the model for the system, the complete optimum linear filter is defined. The accelerometer pulse count registers will be sampled at constant rates. The estimated state vector is extrapolated between measurements according to

$$\hat{\vec{x}}' = \Phi\hat{\vec{x}} + \vec{c} , \tag{15}$$

and the covariance matrix according to

$$E' = \Phi E \Phi^T + S , \tag{16}$$

where Φ and S are pre-computed constant matrices for the time step between measurements. They satisfy the following differential equations:

$$dS/dt = FS + SF^T + N , \qquad S(0) = 0 , \tag{17}$$

$$d\Phi/dt = F\Phi , \qquad \Phi(0) = I , \tag{18}$$

which may be integrated on a digital computer for a time step between measurements of 1 sec (Ref.12). At the time of a measurement, $\hat{x}$ and E are changed according to Equation (4). (See Figure 7 for H , U , and N.)

7. FILTER DESIGN

7.1 Computer Simulations

A complete nonlinear simulation of the inertial platform in a swaying launch vehicle was made on a digital computer in order to simulate real accelerometer outputs. The initial misalignments were 1 degree on all axes; drifts were 10 meru for the vertical and South gyros and zero for the East gyro. The r.m.s. sway in both horizontal directions was 10 cm. The initial conditions for the covariance matrix were 1 deg^2 for the alignment angles and 100 meru^2 for the gyro drifts. All initial cross-correlation terms were assumed zero. The initial estimate of the state was a zero vector.

The response of the filter was excellent for these conditions. A number of runs was first made in which the matrix U was varied so as to cause good agreement between the r.m.s. error as determined by the filter and the actual error. The errors between the estimates and the actual values of azimuth angle, vertical gyro drift and South gyro drift are shown in Figure 8. The errors reach small values for the three cases in 15, 40 and 10 min respectively. The errors in the estimates of the two leveling angles (β and γ) are negligible after the first few measurements. The estimates of the sway variables are not particularly good, but this is not important. These simulations were run under assumed perfect knowledge of East gyro drift because of the classical result that East gyro drift cannot be identified from azimuth error. The fact that East gyro drift must be known presents no problem; as seen from Figure 8, the South gyro can be calibrated in about 10 minutes and the error shows little sensitivity to d_z. This gyro can then be placed East and a complete calibration and alignment made. (The question of other platform positions is discussed later.)

7.2 Design of the Simplified System

The gains for the optimum filter may be pre-computed for all trials since the measurement times will be the same and the *a priori* assumption for the statistics of the initial state vector and noises will not change. For the problem at hand, the implementation of the gains into the system involved, first, the design of a simplified optimum filter. The gains for each state variable depend on both accelerometer measurements and, in general, one gain is much smaller than the other and can be neglected. In this problem all cross-coupling measurement gains are neglected; e.g., vertical drift estimation depends primarily on the South accelerometer so the East accelerometer measurement gain for vertical drift is not implemented. Typically the predominant gains vary as in Figure 9. These gains can be approximated by exponentials and straight-line segments where, at distinct intervals, the time constants and slopes are changed to continually fit the approximate gains to the true gains. The gains for the six sway variables quickly reach steady-state values and may be approximated by three constants.

The response of a simplified filter is shown in Figure 8. The process of design enters since it required a number of runs using different slopes and time constants to get a good match with the response of the complete filter. In fact, in the end, the same exponential gains were used for both po_s and po_e; the same exponential gains were used for β and γ. The total pre-computed constants were 3 sway variable gains, 2 initial conditions for exponentials, and sets of the following 5 numbers which are changed at 10 discrete times: 2 time constants for exponentials (po_s and β) and 3 slopes for straight-line segments (α, d_x, d_y).

7.3 Implementation

A slight variation of this simplified filter was implemented in the Apollo Guidance Navigation and Control system. Part of the program was concerned with initialization for platform positions other than the one considered here. The optimum filter, once implemented, does not change for other platform positions; the measurements that the filter gets are made to simulate the standard platform configuration. For example, if the platform axes were vertical, North and East, and if the sign of the North accelerometer output were changed to simulate a South accelerometer, the filter output for the variables in the North direction need only be interpreted as negative of their true values. Similarly, for some platform positions it is necessary to resolve the measurements to simulate South and East accelerometers.

For some applications, it may be desirable to torque the South gyro at negative horizontal earth rate. The form of the filter and the filter gains do not change because perfect torquing is assumed; just add negative horizontal earth rate to the extrapolation of the angle β. It has also been found convenient to extrapolate the alignment angles according to simple first order ($\alpha = \alpha + \dot{\alpha}dt$, etc.). The sway variables are extrapolated according to a sway transition matrix whose elements can be changed to compensate for variations in launch vehicle parameters.

Once the optimum filter has been implemented according to the simple method outlined, it can be readily adapted to various problems of alignment and calibration. For example, consider the following system test procedure in which the platform axes are identified as x, y, z: (1) run a 10-min test with x up, y South, z East to determine y gyro bias drift; (2) at 10 min read out y drift and use the angle estimates to align the platform; continue to torque the platform at earth rate for 90 sec while counting pulses from the x accelerometer (vertical); (3) orient the platform to x down, y East, z South and run a 10 min test to determine z bias drift; (4) use the angle estimates at 10 min to align the platform and then torque for 90 sec at earth rate while counting x accelerometer pulses; and (5) torque the South gyro at horizontal earth rate for 45 min more while determining vertical drift (x gyro). The y bias as determined in (1) is d_z for this run.

This procedure takes about 68 min, after which enough information is available to determine y gyro bias drift, z gyro bias, the sum of x gyro bias and acceleration sensitive drift, and x accelerometer bias and scale factor. One can readily imagine how an automated system test procedure can be set up to completely calibrate the system in the swaying spacecraft. The last step in the program would be an alignment run to ready the system for launch. And through all of this, the basic simplified optimum filter does not change.

7.4 Hardware Problems

The primary source of azimuth error is the uncertainty in East gyro drift which comes about from errors in the calibration of the gyro; the primary source of vertical drift error is due to variations in the East gyro drift during a test. If the East gyro has a large drift due to acceleration along its input axis, then it is desirable to keep the input axis almost horizontal during the test by torquing the South gyro at negative earth rate to minimize the variations in East gyro drift. Unfortunately, if the bias of the South gyro changes when it is torqued, then the South gyro calibration will yield two answers, corresponding to the torqued and untorqued cases, respectively. (Torquing was done in the system test program.)

Another possible problem area appears when using pulsed integrating accelerometers. If either horizontal accelerometer has a large dead-zone for near zero inputs, large transients in the filter output will appear. The form of the transients will vary, depending on the time during the test that the accelerometer goes through the dead-zone. Vertical drift estimation is particularly sensitive to a dead-zone in the South accelerometer. In some cases a transient on the order of 800 meru has been observed; the filter never reached the correct value of drift at the end of 45 minutes because the vertical drift gain is small at the end of the test. As a practical solution to the problem, the platform is deliberately offset (between steps (4) and (5) in the system test program) from the vertical before beginning a vertical drift test so that the accelerometer never goes through null. This problem is indicative of the strange results that can occur when the model for the system is incorrect. Philosophically, we have designed a total system test and as such it should indicate in some manner out-of-specification conditions which would then require lower level testing. The model must therefore include all in-spec conditions and the designer's experience must be used to recognize out-of-spec situations.

7.5 Laboratory Test Results

As part of the filter implementation verification, the Apollo inertial platform was suspended by a cable from the laboratory ceiling. The platform was pushed to simulate spacecraft motion on the launch pad. Comparisons with independent measurement techniques verified the program accuracy to within the limits of the gyro performance.

8. CONCLUSION

The possibility exists for significant reduction in the errors in calibration systems by using statistical estimation techniques that make the most efficient use of all available information. It also appears likely that suitable simplifications can be made to adapt these techniques to practical computers, as was illustrated with two examples.

ACKNOWLEDGEMENT

This report was prepared under DSR Project 55-23870, sponsored by the Manned Spacecraft Center of the National Aeronautics and Space Administration through Contract NAS 9-4065 with the Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

REFERENCES

1. *Optimum Utilization of Information in SINS*. Reports R-24 and R-26, Dynamics Research, Stoneham, Mass., 1963.

2. Fagen, S.L. *A Unified Approach to the Error Analysis of Augmented Dynamically Exact Inertial Navigation Systems*. Institute of Electrical and Electronic Engineers, Transactions on Aerospace and Navigation Electronics, Vol.ANE-11, No.4, December 1964, pp.234/248.

3. Brown, R.G.
Priest, D.T. *Optimization of a Hybrid Inertial Solar-Tracker Navigation System*. Institute of Electrical and Electronic Engineers, International Convention Record, Vol.12, Pt.7, 1964, pp.121-135.

4. *Theoretical Foundation of AKURON*. Autonetics Division, North American Aviation, Inc.

5. Astrom, K.J. *Some Problems of Optimal Control in Inertial Guidance*. Research Paper RJ-229, IBM Research Laboratory, San Jose, California, November 1962.

6. Astrom, K.J. *On the Problem of Indicating the Vertical in a Moving Vehicle*. Foa 2 Rapport A 2306-233, Forsvarets-Forskningsanstalt, Avdelning 2, Stockholm 80, Sweden, October 1964.

7. Kalman, R.E. *A New Approach to Linear Filtering and Prediction Problems*. Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers, Vol.82, March 1960, pp.35-45.

8. Kalman, R.E.
Bucy, R.S. *New Results in Linear Filtering and Prediction Theory*. Journal of Basic Engineering, Transactions of the American Society of Mechanical Engineers, Vol.83, March 1961, pp.95-108.

9. Potter, J.E. *A Guidance-Navigation Separation Theorem*. Report RE-11, Massachusetts Institute of Technology, Experimental Astronomy Laboratory, Cambridge, Mass., August 1964.

10. Brock, L.D. — *Application of Statistical Estimation to Navigation Systems.* Ph.D. thesis, T-414, Department of Aeronautics and Astronautics, Massachusetts Institute of Technology, June 1965.

11. Schmidt, G.T. — *System Test Determination of Apollo IRIG Drift Coefficients.* Massachusetts Institute of Technology, Instrumentation Laboratory Report T-421, June 1965.

12. Schmidt, G.T. — *Apollo System Test Group Memo No.459.* Massachusetts Institute of Technology, Instrumentation Laboratory, June 1965.

13. James, H.M. et al. — *Theory of Servomechanisms.* Dover Publications, New York, 1965, p.369.

14. Markey, W. Hovorka, J. — *The Mechanics of Inertial Position and Heading Indication.* John Wiley, New York, 1961.

15. Bona, B.E. Sinay, R.J. — *Optimum Reset of Ship's Inertial Navigation System.* Institute of Electrical and Electronic Engineers, Transactions on Aerospace and Electronic Systems. Vol.AES-2, No.4, July 1966, pp.409-414.

16. Dusek, H. — *Theory of Error Propagation in Astro-Inertial Guidance Systems for Low-Thrust Earth Orbital Missions.* Preprint No.2682-62, American Rocket Society, Nov.1962.

17. Bona, B.E. Hutchinson, C.E. — *Optimum Reset of an Inertial Navigator from Satellite Observations.* Proceedings of the National Electronics Conference, Chicago, Ill., October 1965.

18. Denhard, W.G. — *Laboratory Testing of a Floated, Single-Degree-of-Freedom, Integrating, Inertial Gyro.* Report R-105, Massachusetts Institute of Technology, Instrumentation Laboratory, September 1956.

19. — *Updated Model Evaluation for Polaris SINS Gyros.* Report E-458, Dynamics Research Corporation, Stoneham, Mass., Nov. 1964.

20. Schmidt, G.T. Brock, L.D. — *Statistical Estimation in Inertial Navigation Systems.* AIAA Journal of Spacecraft and Rockets, Vol.5, No.2, Feb. 1968, pp.146-154.

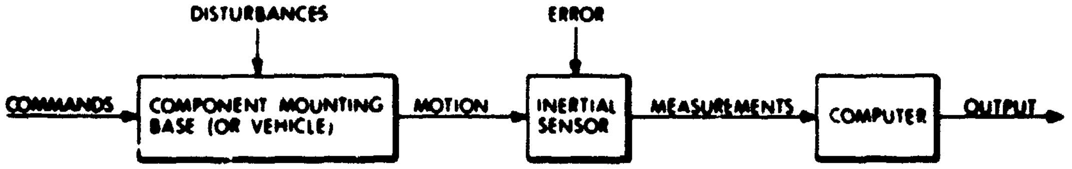


Fig.1 Functional form of calibration system

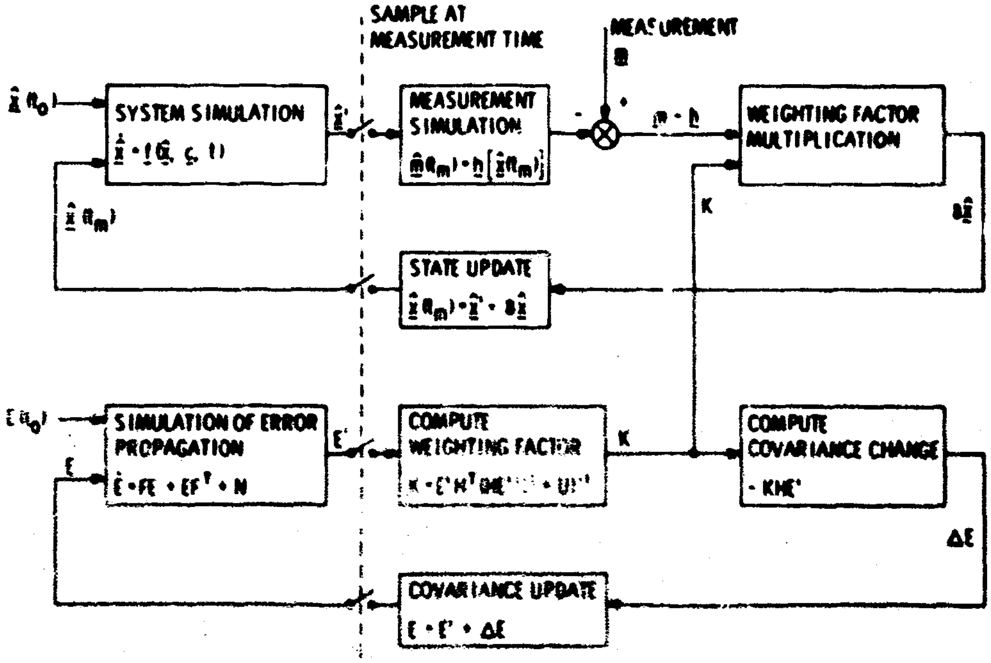

SAMPLE AT MEASUREMENT TIME
MEASUREMENT
$\underline{m}$
$\hat{\underline{x}}(t_0)$
SYSTEM SIMULATION
$\dot{\hat{\underline{x}}} = \underline{f}(\hat{\underline{x}}, \underline{c}, t)$
$\hat{\underline{x}}'$
MEASUREMENT SIMULATION
$\hat{\underline{m}}(t_m) = \underline{h}[\hat{\underline{x}}(t_m)]$
-
+
$\underline{m} - \underline{h}$
WEIGHTING FACTOR MULTIPLICATION
$\hat{\underline{x}}(t_m)$
K
$\delta\hat{\underline{x}}$
STATE UPDATE
$\hat{\underline{x}}(t_m) = \hat{\underline{x}}' + \delta\hat{\underline{x}}$
$E(t_0)$
SIMULATION OF ERROR PROPAGATION
E
$\dot{E} = FE + EF^T + N$
E'
COMPUTE WEIGHTING FACTOR
K
COMPUTE COVARIANCE CHANGE
$-KHE'$
ΔE
COVARIANCE UPDATE
$E = E' + \Delta E$

Fig.2 Optimum filter functional diagram

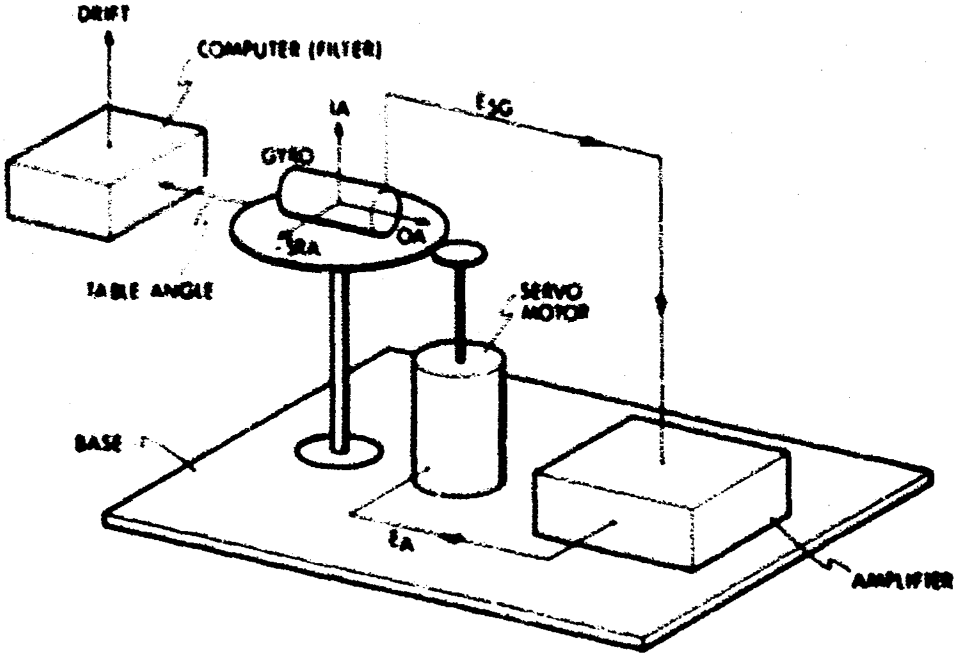


Fig.3 Gyro servo test calibration

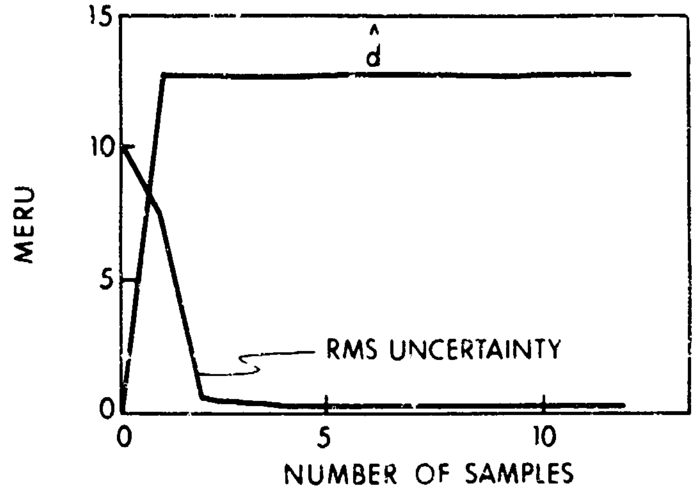


Fig.4 $\hat{d}$ and r.m.s. uncertainty

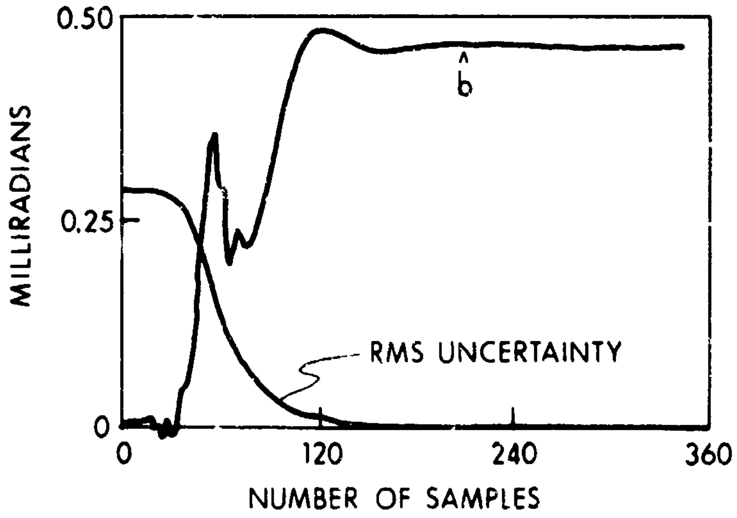


Fig.5 $\hat{b}$ and r.m.s. uncertainty

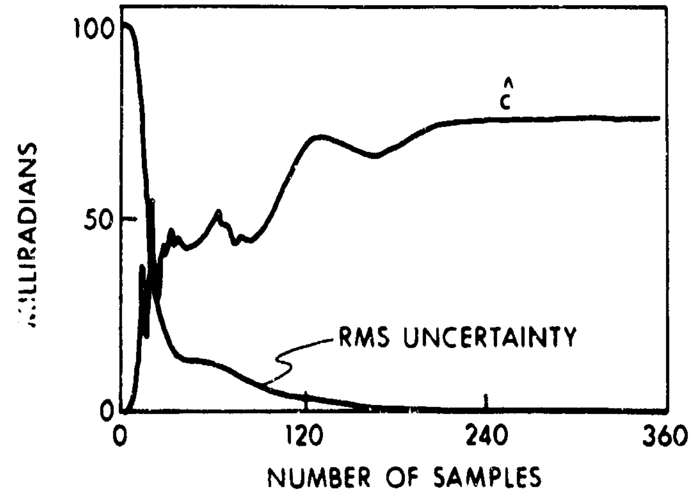


Fig.6 $\hat{c}$ and r.m.s. uncertainty

$$\underline{x} = \begin{bmatrix} po_s \\ po_e \\ v_s \\ v_e \\ a_s \\ p_s \\ a_e \\ p_e \\ \alpha \\ \beta \\ \gamma \\ d_x \\ d_y \end{bmatrix} \quad \cdot\underline{c} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ t_x - \Omega_v \\ t_y + \Omega_h \\ t_z + d_z \\ 0 \\ 0 \end{bmatrix} \quad \underline{n} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ n \\ 0 \\ n \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix}$$

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -g & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & f_1 & 0 & f_2 & f_3 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & f_1 & 0 & 0 & f_2 & f_3 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Omega_h & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \Omega_v & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & -\Omega_h & -\Omega_v & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$f_1 = -\omega_n^2 - 2\zeta\lambda\omega_n$$

$$f_2 = -\lambda - 2\zeta\omega_n$$

$$f_3 = -\lambda\omega_n^2$$

$$U = \begin{bmatrix} ra & 0 \\ 0 & ra \end{bmatrix} \quad H = \begin{bmatrix} 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

N = 13X13 ZERO MATRIX EXCEPT FOR ROW 5, COLUMN 5 AND ROW 7, COLUMN 7 WHICH ARE BOTH nw

Fig.7 Filter matrices and vectors

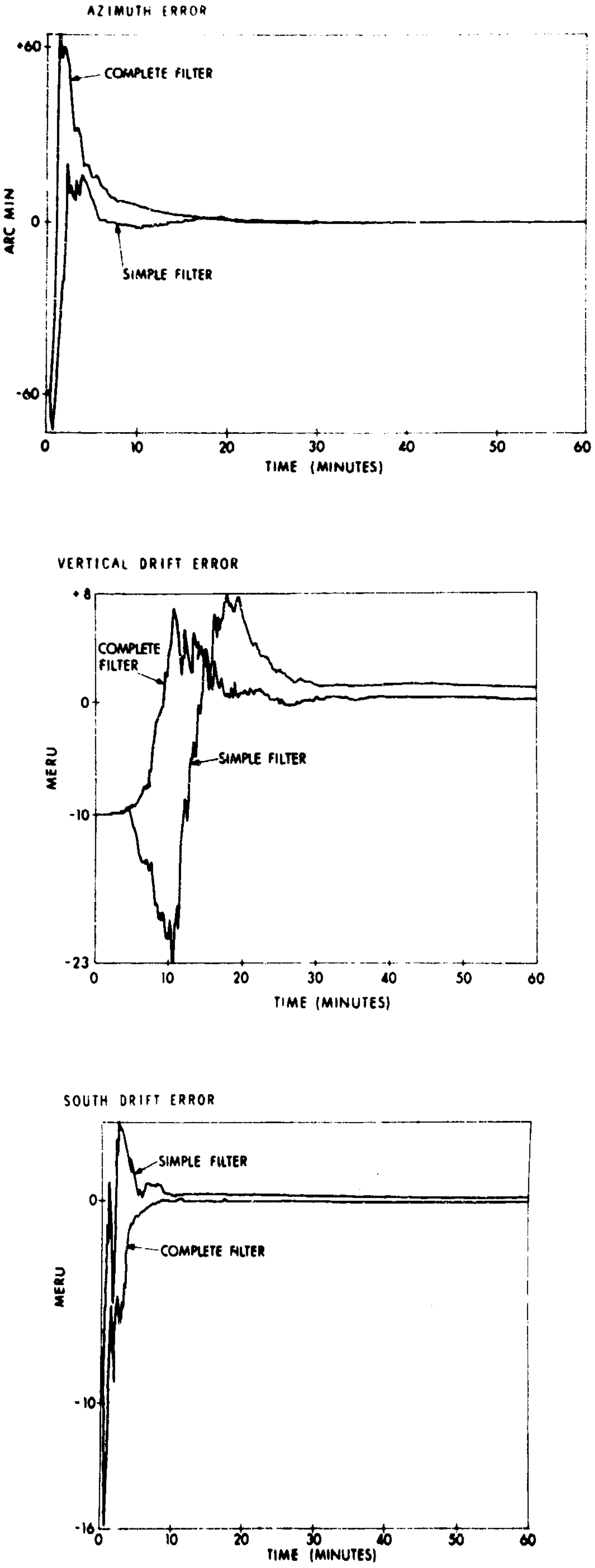


Fig.8 Complete and simplified filter performance

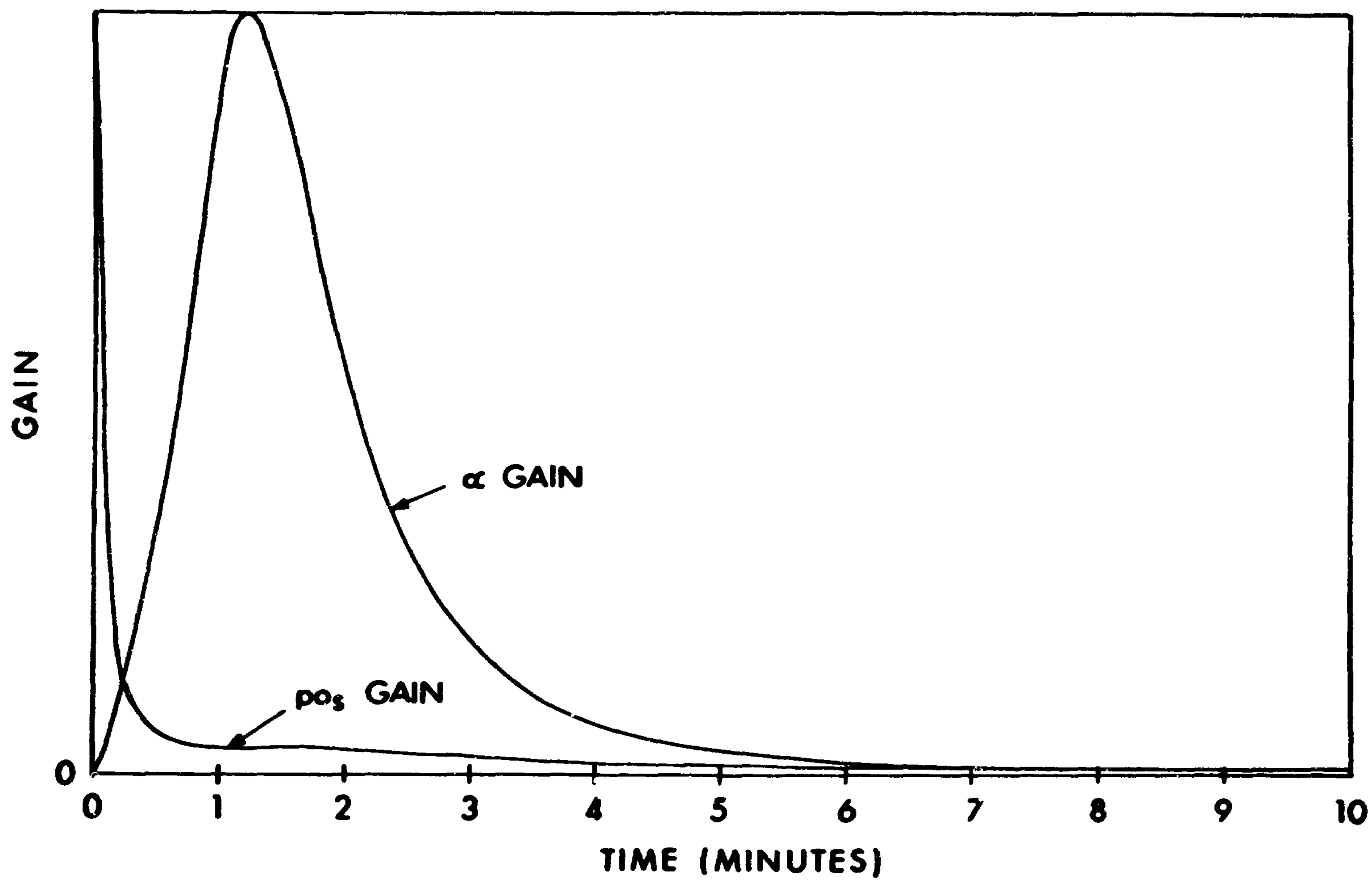


Fig.9 Typical gain histories

EULER ANGLE STRAPPED-DOWN COMPUTER

by

Alan van Bronkhorst

Lear Siegler, Inc.,
Santa Monica, California, USA

EULER ANGLE STRAPPED-DOWN COMPUTER

Alan van Bronkhorst

Navigation - it's a magic word, the key to travel and adventure. We covet and ponder with awe the innate navigation ability of other species of living things. A pigeon, released at a strange location, climbs swiftly, circles, and with wings flashing white and grey, sets his course for home. Lacking instinct, we are forced to rely on our intellect for navigation. With what results? First, we are often lost, and second, we have come to set great store by development of electromechanical "black boxes" which perform navigation for us.

This second result is particularly true if the "black boxes" meet military specifications: smaller size, lower cost, improved performance, and better reliability than currently available equipment. The present flurry of activity in development of inertial strapped-down systems is engendered by the hope of achieving at least the first two of these four goals (which is not a bad average) and is promoted by the emerging availability of the necessary high-speed computers.

A particular type of whole-number high-speed computer, solving a three-parameter Euler angle transformation, has been investigated as a component of a strapped-down mid-course inertial guidance system for a short range missile. This paper presents the results of this investigation.

What are the inherent advantages of a whole-number digital computer for a strapped-down inertial system?

Any strapped-down inertial system computer must, at the very least, convert the outputs of orthogonal body-fixed accelerometers into velocity vectors in some navigation coordinate system. Typically, an incremental computer (DDA) performs this conversion with a 3 × 3 direction cosine matrix that it generates from outputs of orthogonal body-fixed rate gyros. This type of strapped-down inertial measurement unit has been previously investigated and reported in detail by others[1-4]. The literature abounds with computer algorithm optimization studies. A total guidance system, however, needs not just an inertial measurement unit but additional computer functions as well. These functions must include solution of the navigation and guidance equations and, preferably, such niceties as sensor corrections, system self-check, auxiliary guidance system integration and even gain and frequency compensation of control system commands. Using a strapped-down inertial measurement unit (IMU) with outputs from an incremental computer, the guidance system design has two alternatives:

(i) Expand the DDA to perform the total system requirements, or

(ii) Supplement the DDA conversion computer with a separate whole-number guidance computer.

Neither of these solutions is very attractive on an overall system basis.

J.C.Wilcox, of TRW Systems, in a recent paper[5], writes

> "The incremental computer is subject to the constraint that the computation rate must be greater than or equal to the maximum input rate divided by the quantum size. In order to maintain adequate resolution, all calculations must proceed

at the same high rate. The addition of new calculations exacts penalties in both computer speed and storage requirements. This constraint does not apply to the whole number computer.

In the whole number computer, the penalty for additional calculations may be only in computer storage requirements, ..."

A more flexible strapped-down guidance system, then, is an inertial sensor package (gyros and accelerometers) combined with a single whole-number computer which can perform both coordinate conversion and the navigation/guidance functions simultaneously. This flexibility results from decreased costs and simplified interface with related subsystems.

Although a whole-number computer can be programmed to solve the direction cosine strapped-down transformation, this study investigated a three-parameter Euler angle transformation technique. The equations are derived by simply relating gimbal angle rates to body-axis rates by transformation through the equivalent gimbal angles. Equation (1) represents a conventionally gimballed three-axis platform.

$$\begin{bmatrix} p \\ q \\ r \end{bmatrix} = \begin{bmatrix} \dot{\phi} \\ 0 \\ 0 \end{bmatrix} + [\phi] \begin{bmatrix} 0 \\ \dot{\theta} \\ 0 \end{bmatrix} + [\phi]\ [\theta] \begin{bmatrix} 0 \\ 0 \\ \dot{\psi} \end{bmatrix} \qquad (1)$$

where p, q, r are body-axis rates and $[\phi]$, $[\theta]$ are rotational matrices of roll and pitch angles.

These equations are transformed to

$$\begin{aligned} \dot{\theta} &= -r \sin \phi + q \cos \phi \\ \dot{\psi} &= (r \cos \phi + q \sin \phi) \sec \theta \\ \dot{\phi} &= p + \dot{\psi} \sin \theta \end{aligned} \qquad (2)$$

and, finally, velocities are transformed from body-axes to inertial coordinates by sequential rotation through the three generated angles:

$$\begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix}_i = [\psi]\,[\theta]\,[\phi] \begin{bmatrix} \dot{x} \\ \dot{y} \\ \dot{z} \end{bmatrix}_{b.\,a.} \qquad (3)$$

The reasons for investigating this type of transformation are:

(i) Lear Siegler Inc., has developed previously - for other navigation systems - a whole-number, vector computer[6,7]. The vector computer has an arithmetic unit which is organized for rapid shift and add processing of vector operations such as rotation and resolution. This type of computer is ideally suited for the straightforward Euler angle coordinate conversion process.

(ii) The transformation, by its nature, is always essentially orthonormal and provides simplified updating for integrated navigation systems applications.

(iii) Attitude outputs are directly available for vehicle control/display, trajectory shaping, system initialization and sensor compensations.

Is this transformation technique usable? In spite of the claims we have made, the answer, of course, depends on the cost of the computer to achieve adequate performance. We now examine the errors as they affect computer software.

The errors of strapped-down system computers come from inaccurate processing of the outputs of the inertial sensors. These errors accrue both in generating the transformation matrix and in applying the transformation to the measured velocity vectors. We may classify these errors as orthogonality errors, angularity errors, and normality errors. By analogy with gimbal coordinate converter resolvers (whose function the conversion matrix duplicates) these errors are inter-axis errors, gimbal drift errors, and transformation ratio errors, respectively.

Clearly the inter-axis, or orthogonality, errors of the matrix are zero. This results because we are generating and computing with the angles directly, and not their direction cosines. We are concerned then only with velocity vector angle (or direction) transformation errors which are the result of gimbal drift, and with velocity vector magnitude transformation errors which arise, even though the transformation matrix itself is nearly normal, from other computational processes.

The sources of these errors, as in any strapped-down system, are quantization, commutation, truncation, round-off, and conversion. The order of listing here is not significant except as an index for the following discussion.

Quantization, or sensor storage error, results from the process of converting analog quantities measured by the sensors to discrete angle and velocity input increments to the computer. Because the computer is synchronous, it accepts input pulses for operations only at the start of each iteration cycle. As a result, it is frequently operating on information that lags the true sensor output. For example, consider a gyro with an output increment pulse size of 0.03 degrees, an input rate of 20 degrees per second and supplying a computer with an iteration time of one millisecond. Table I shows the quantization error.

From this table we can guess that the average quantization error is on the order of one-half the increment size, h . Why not just make h very small? Well, h is the maximum input (angular rate of linear acceleration) divided by the maximum pulse repetition frequency (PRF) of the analog-to-digital converter. The guidance system designer can do little about the maximum inputs of the vehicle dynamics except complain to the project aerodynamicist, who probably couldn't care less. And in increasing the PRF we soon start paying for decreasing quantization errors with increasing non-linearity errors which result from the A/D conversion process, and increasing input hardware requirements. For this system, synthesized for mid-course guidance of a short range missile, the inputs are body-axis rates of 300 degrees per second, and linear accelerations of 20 g. Our error crossover point, determined empirically, is at PRF $\simeq$ 10 kpps. This results in inputs of approximately 0.03 degrees per pulse or 0.06 feet per second per pulse. The precise values are set by different criteria, as shown later.

The second definable source of errors is in commutation. Commutation errors are both angle and velocity. The former come principally from the non-commutative properties of angular increments under vector operations. The error limits may be established quite simply. The difference equations in the solution sequence are:

$$\left.\begin{array}{lll}
1. & \Delta\theta_N = (-r\Delta t)\sin\phi_{N-1} + (q\Delta t)\cos\phi_{N-1} & \\
 & \theta_N = \theta_{N-1} + \Delta\theta_N & \\
2. & \Delta\psi_N = (r\Delta t)\cos\phi_{N-1} + (q\Delta t)\sin\phi_{N-1}/\cos\theta_N & \\
 & \psi_N = \psi_{N-1} + \Delta\psi_N &
\end{array}\right\} \qquad (4)$$

$$3. \quad \left. \begin{aligned} \Delta\phi_N &= (p\Delta t) + \Delta\psi_N \sin\theta_N \\ \phi_N &= \phi_{N-1} + \Delta\phi_N \,. \end{aligned} \right\} \qquad (4)^*$$

Assume an input rate $\vec{\omega} + \vec{i}p + \vec{j}q + \vec{k}r$, where $p = q = r$.

For Iteration 1, $\vec{\omega}_1 = \vec{\omega}$.

For Iteration 2, $\vec{\omega}_2 = -\vec{\omega}$,

Let

$$p\Delta t = q\Delta t = r\Delta t = \Delta\omega$$

$$\phi_{N-1} = \theta_{N-1} = \psi_{N-1} = 0^0 \,.$$

If we turn the crank through two iterations, using these inputs,

$$\theta_{N+1} = \phi_{N+1} = \psi_{N+1} = 0^0 \pm \text{commutation errors.}$$

These errors turn out to be on the order of $\Delta\bar{\omega}^2/2$ per iteration. Please observe that this error results from angular vibration inputs about axes which are not coincident with the principal body-axes. Although similar in form, it should not be confused with 1:1 limit cycle error which is frequently analyzed in the DDA/direction cosine literature. A 1:1 limit cycle error is a truncation, or computation, error which results from angular vibration inputs about any axis because the incremental computer mechanizes only an approximation to the true direction cosine equation solution.

Velocity/angle commutation error results because, although the angles are constantly changing during an iteration period in which a velocity increment is measured, this entire velocity increment is converted through the final value angle only.

The procedure for each iteration period is:

1. Measure angle and velocity increments
2. Update the angle values.
3. Use these updated angles to transform the measured velocity increments.

For example, the transformation matrix is

$$\begin{bmatrix} x \\ y \\ x \end{bmatrix}_i \begin{matrix} = \\ = \\ \end{matrix} \begin{bmatrix} c\phi\, c\theta\, c\psi - s\phi\, s\psi & s\phi\, c\theta\, c\psi + c\phi\, s\psi & -s\theta\, s\psi \\ -c\phi\, c\theta\, s\psi - s\phi\, c\psi & -s\phi\, c\theta\, s\psi \quad c\phi\, c\psi & s\theta\, s\psi \\ c\phi\, s\theta & -s\phi\, s\theta & c\theta \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \end{bmatrix}_{b.a.} \qquad (5)$$

where s and c are the sine and cosine of the respective Euler angles.

Assume that $\phi_0 = \theta_0 = \psi_0 = 0$ and p, q, r , are constants such that, after one iteration, $\phi = \theta = \psi = \Delta\omega$, where $\Delta\omega$ is a small angle so that $c\Delta\omega \simeq 1.0$ and $s\Delta\omega = \Delta\omega$. Now if $\overline{\Delta X}_1\, \overline{\Delta Y}_1\, \overline{\Delta Z}_1$, are increments of velocity accumulated during the first iteration period, then

---

*Equation (4) is slightly time-skewed to conserve computer memory
Solution with θ_{N-1} instead of θ_N has no appreciable effect on accuracy

$$\begin{bmatrix} \dot{\overline{\Delta X}}_1 \\ \dot{\overline{\Delta Y}}_1 \\ \dot{\overline{\Delta Z}}_1 \end{bmatrix}_i = \begin{bmatrix} 1-\overline{\Delta\omega}^2 & 2\Delta\omega & -\Delta\omega \\ -2\Delta\omega & 1-\Delta\omega^2 & \overline{\Delta\omega}^2 \\ \Delta\omega & -\overline{\Delta\omega}^2 & 1 \end{bmatrix} \begin{bmatrix} \dot{\overline{\Delta X}}_1 \\ \dot{\overline{\Delta Y}}_1 \\ \dot{\overline{\Delta Z}}_1 \end{bmatrix}_{b.a.} \tag{6}$$

and if p_2, q_2, r_2, are approximately $-p_1, -q_1, -r_1$, such that, after the second iteration, $\phi_2 = \theta_2 = \psi_2 = 0$, and if $\dot{\overline{\Delta X}}_2, \dot{\overline{\Delta Y}}_2, \dot{\overline{\Delta Z}}_2 = -\dot{\overline{\Delta X}}_1, -\dot{\overline{\Delta Y}}_1, -\dot{\overline{\Delta Z}}_1$ then

$$\begin{bmatrix} \dot{\overline{\Delta X}}_2 \\ \dot{\overline{\Delta Y}}_2 \\ \dot{\overline{\Delta Z}}_2 \end{bmatrix}_i = [I] \begin{bmatrix} \dot{\overline{\Delta X}}_2 \\ \dot{\overline{\Delta Y}}_2 \\ \dot{\overline{\Delta Z}}_2 \end{bmatrix}_{b.a.} = [I] \begin{bmatrix} -\dot{\overline{\Delta X}}_1 \\ -\dot{\overline{\Delta Y}}_1 \\ -\dot{\overline{\Delta Z}}_1 \end{bmatrix}_{b.a.} \tag{7}$$

$$\left.\begin{aligned} \dot{\overline{\Delta X}}_1 + \dot{\overline{\Delta X}}_2 &= -\Delta\omega\,(2\dot{\overline{\Delta Y}}_1 - \dot{\overline{\Delta Z}}_1) \\ \dot{\overline{\Delta Y}}_1 + \dot{\overline{\Delta Y}}_2 &= 2\Delta\omega\,(\dot{\overline{\Delta X}}_1) \\ \dot{\overline{\Delta Z}}_1 + \dot{\overline{\Delta Z}}_2 &= \Delta\omega\dot{\overline{\Delta X}}_1\ . \end{aligned}\right\} \tag{8}$$

Then the error per iteration can be on the order of $\frac{1}{2}\Delta\omega\ \dot{\overline{\Delta X}}_1$. Obviously commutation error will decrease with increasing computer iteration rates. Frank B.Hills has a detailed discussion of velocity/angle commutation errors in Reference 3.

Truncation errors, the third source on our classification list, result primarily from digital summation of incremental quantities rather than true integration. For evaluation of the maximum bound of angle truncation errors, we expand Equation 2(a), as typical, in a Taylor series,

$$f(\phi) = \dot{\theta} = -r \sin \phi + q \cos \phi\ . \tag{9}$$

then

$$f'(\phi) = -\dot{r} \sin \phi - q\dot{\phi} \sin \phi + \dot{q} \cos \phi - r\dot{\phi} \cos \phi \tag{10}$$

and, assuming constant body rates for one iteration period,

$$f'(\phi) = -q\dot{\phi} \sin \phi - r\dot{\phi} \cos \phi \simeq r\dot{\phi}\ (\text{maximum})\ . \tag{11}$$

The simplest type of digital integration is rectangular, or first-order. For this method, the error per iteration is

$$|Ex|_1 = (r\Delta t)(\dot{\phi}\Delta t)(\tfrac{1}{2}) \simeq \overline{\Delta\omega}^2$$

Truncation errors may be reduced by either increasing computer iteration rates or by increasing computer functions and, unavoidably, computer hardware to obtain derivatives and implement higher order integration. For example, for second-order integration, the errors per iteration become,

$$|Ex|_2 \simeq \frac{\dot{\overline{\Delta\omega}}^3}{2}\ .$$

There is, of course, a velocity truncation error also, but this error is the same for either strapped-down or gimbaled inertial systems and will not be considered in this paper. Truncation errors for various digital computer algorithms for updating the direction cosine transformation matrix are discussed by W.F.Ball[2]. (Also see References 1, 3 and 4).

Round-off, or resolution, errors are the result of the finiteness of digital representation of quantities that are generated during the computation processes. Any number in the computer may be represented only within $\pm \frac{1}{2}$ LSB (least significant bit), and these errors may accumulate with large numbers of computations. The obvious way to keep round-off errors insignificant is to increase word lengths. But long word lengths require more computer hardware and decrease computer iteration rates which, sadly, increases the truncation and commutation errors. A compromise solution is to increase word lengths for only certain especially sensitive computations, such as DIVIDE, by float operations. Round-off is also precisely and importantly controlled at the input by making each increment pulse an integral number of LSB's. It is this constraint that accurately specifies pulse size, and simultaneously provides a convenient method of compensating for sensor errors.

Finally, conversion error is simply the effect, or error, in computing the inertial velocity vector as a function of the errors in the transformation matrix.

Consider velocity vector increments under a single rotational transformation,

$$\begin{bmatrix} \dot{\overline{\Delta x}} \\ \dot{\overline{\Delta y}} \\ \dot{\overline{\Delta z}} \end{bmatrix}_{I} = [1 \pm \delta] \begin{bmatrix} 1 & 0 & 0 \\ 0 & c(\Phi \pm \epsilon\Phi) & s(\Phi \pm \epsilon\Phi) \\ 0 & -s(\Phi \pm \epsilon\Phi) & c(\Phi \pm \epsilon\Phi) \end{bmatrix} \begin{bmatrix} \dot{\overline{\Delta x}} \\ \dot{\overline{\Delta y}} \\ \dot{\overline{\Delta z}} \end{bmatrix}_{b.a.} \qquad (12)$$

where δ is the transformation ratio error resulting from round-off, and $\epsilon\Phi$ is the angularity error resulting from quantization, angle commutation, and truncation. Expanding the sine, cosine terms and substituting, it may be shown that the errors are on the order of

$$Te \simeq (\delta \pm \epsilon\Phi)\dot{\overline{\Delta x}}$$

per rotation. If the principal angle errors are independent, then

$$Te \simeq \sqrt{(2)}(\delta \pm \epsilon\Phi)\dot{\overline{\Delta x}}$$

for each transformation, since each velocity increment is rotated only through two angles.

The conclusion of this simple error study is identical with the results of the most elegant analysis: the output errors of any strapped-down system computer are a function of both its sophistication and the input information which it must process. And for inputs which approach realism, mathematical error modeling becomes impossible.

Simulation is a sophisticated and essential tool for computer synthesis. For any inputs, word lengths, algorithms, iteration rates, and quantization levels may be traded to obtain the required performance at minimum cost. The simulation technique block diagram is shown in Figure 1. The reference solution processed the input data without quantization, at an iteration rate of 2000 times per second using Runge-Kutta fourth-order integration schemes and double precision, 64-bit word length for all calculations. The vector computer simulation forced our IBM-360 machine to calculate with special vector algorithms using a two's-complement binary number system.

For computer synthesis, simple inputs are of little value. Principally they are esteemed by those who appreciate the esthetics of verifying empirically the error analysis for simple inputs. In fact, F.B.Hills^{3} states

> "Computer simulation tests in which the attitude changes are rotation about one axis, vibrations about one axis, or constant rotations about three axes should show good results, for these are rotations for which the components of ω are proportional, which the algorithms assume.....Exclusive use of these tests can yield misleading results concerning the necessary data processing rates."

For realistic inputs to synthesize a strapped-down inertial guidance computer for a typical missile, we used the dynamics of a typical missile which was simulated by the Martin Company, Orlando, Florida. Two trajectories of an air-to-surface missile, fired at targets 25 nautical miles behind and 50 nautical miles ahead of the launch vehicle, were supplied. This information was in the form of computer print-out showing time, body-axis angular rates and angular accelerations, and linear accelerations and position in an inertial coordinate system.

The data start at time of booster ignition, $t = L + 1.0$ seconds, where L is launch time. For the first 1.0 seconds of flight, we synthesized an initial trajectory from the following parameters:

(a) The missile is launched by firing downward with an impulse of 12.5 g for 50 msec. Forward velocity is almost constant at 0.5 Mach and vertical velocity at $t = L + 1.0$ is 20 ft/sec.

(b) $L < t \leq L + 1.0$ seconds the missile roll stabilizes. The roll stabilization system is first-order. Maximum roll amplitude may be up to 90 degrees, and maximum roll rate is 300 degrees per second.

(c) Yaw and pitch oscillation occurs during the first second at ± 10 degrees amplitude, approximately 0.5 Hz, $\dot{\theta}_{max} = \dot{\psi}_{max} = 30$ degrees per second.

(d) First body-bending frequency is approximately 40 Hz across the longitudinal axis. However, this frequency is not excited until after booster ignition, at $t = L + 1.0$ seconds.

(e) The angular velocities of the body-axes at $t = L + 1.0$ are p, r = zero, $q = -10$ degrees per second.

Under "worst case" assumptions, the analytic form of body-axis rate inputs for the launch phase which satisfied these constraints and boundary values may be determined.

$$\begin{aligned} \phi &= A \cos \omega_1 t = (\pi/2) \cos \pi t + \pi/2 \\ \psi &= A \sin \omega_2 t = 0.175 \sin 3t \qquad (13) \\ \theta &= A \cos (\omega_2 t + \pi/6) = 0.175 \cos (3t + \pi/6) . \end{aligned}$$

These values of ψ, θ, ϕ for $L < t \leq 1.0$ seconds are shown in Figure 2. Straightforward substitution of the values in the Euler rate equations provide the analytic expressions for p, q, and r, the body-axis angular rate inputs.

After booster ignition ($L + 1.0$) a continuous time function of the missile body-axis rates was obtained from the trajectory simulation provided as discrete print-outs by the Martin Company. This function was generated by matching a fifth-order polynomial through three successive discrete printed points for each of the body-axis angular rates. Simultaneously the derivative of this polynomial was matched through three, time corresponding, successive discrete printed points for each of the three body-axis angular accelerations. The solution of these polynomials as functions of time provide body-axis rate inputs to the simulated computer. A similar technique generated an analytic velocity input. These inputs, for the first fifteen seconds of flight, are shown in Figures 3 and 4. The inputs were terminated after 15 seconds, the end of boost phase, to save simulation time and because, after this time, the angular rates and linear accelerations are both very reduced from the initial flight phase.

With a fixed quantization level our trade parameters are word length, iteration rate, and integration algorithm. Figures 5 and 6 show the effect of word length on transformation accuracy for the launch phase input. The iteration rate here was arbitrarily picked at 1 millisecond. For this input 20-bit words are shortest which do not show round-off error degradation. In these figures the errors shown are within the expected quantization level error. In the launch phase velocity errors are negligible because of the short acceleration input time. Figures 7 and 8 show the effect of iteration rate on transformation accuracy for the launch phase input.

Apparently there is no degradation of angle computation even at rates as low as four milliseconds. However, although not shown here, there is an increase in angle/velocity commutation error at low solution rates. We chose one millisecond because, for this computer, high solution rates were not as expensive as long word lengths.

A second-order trapezoidal integration scheme was also tried using one millisecond iteration rates and 20-bit word lengths. Any accuracy improvement was masked by the quantization errors. This agrees with a prediction by T.F.Wiener[1] that an overall improvement may not result from implementing higher order integration because of the magnitude of errors from other sources.

Because the launch phase, though very severe in dynamics, was too short to observe any error growth rates, we provided a moderate missile maneuver of 10 degrees per second turn rate, 5 degrees per second pitch, and a superimposed roll oscillation. This is a three-axis rate input that varies with time in both magnitude and direction and should show error which builds from commutation and truncation. Figure 9 shows this input and the transformation error matrix. There appears a small bias error in ψ from an unknown source, but error growth rates are small compared to the quantization level scatter.

The final test of our synthesized computer are the simulated missile flight trajectories. Angle and position errors as a function of time are shown in Figures 10 and 11. The errors at the end of boost phase, extrapolated to impact, give miss distances of 30 and 100 feet cep. This will not significantly increase the overall missile guidance system design accuracy.

By this time some sharpshooter in the back row is impatient to ask about gimbal lock. Our equivalent of a three-axis gimbal set does not have the mechanical gimbal problem of providing unlimited torques at high pitch angles, but computational accuracy is troublesome. In Equation (2) we are forced to divide by zero as θ approaches 90 degrees: a singularity, in the mathematical idiom. Two general techniques for avoiding this singularity, are a pseudo-four-axes gimbal set or a gimbal shifting scheme.

The rate equations for an equivalent four-axes set are

$$\left.\begin{aligned} \dot{\phi}_1 &= k_2 \sin\theta + (p - \dot{\phi}_0)\cos\theta \\ \dot{\psi} &= [k_1 \cos\theta - (p - \dot{\phi}_0)\sin\theta]/\cos\phi_1 \\ \dot{\theta} &= k_1 - \psi \sin\phi_1 \ , \end{aligned}\right\} \qquad (14)$$

where

ψ = azimuth gimbal angle

ϕ_1 = inner roll gimbal angle

θ = pitch gimbal angle

ϕ_0 = outer roll gimbal

$k_1 = q\cos\phi_0 - r\sin\phi_0$

$k_2 = q\sin\phi_0 + r\cos\phi_0$.

This is a set of three equations with four variables and, furthermore, we have not eliminated our chance to divide by zero. What we need is a restraint equation relating ϕ_0 to ϕ_1, such that $\cos\phi_1 > 0$.

It is possible to convert inner roll gimbal rates to the outer roll axis and implement an effective outer roll rate servo to null inner roll axis rates. This requires more computations in the actual transformation matrix and hence will reduce accuracy.

A simple method of extending the attitude capability is to let $\dot{\phi}_0 = p$. Since p is body-axis roll rate, this should insure that ϕ_0 is normally large with respect to ϕ_1 . The exact ratio, of course, is a function of vehicle flight path. But, with this restraint, the equations become

$$\left.\begin{aligned} \dot{\phi}_0 &= p \\ \dot{\phi}_1 &= k_2 \sin\theta \\ \dot{\psi} &= k_2 \cos\theta \,/\, \cos\phi_1 \\ \dot{\theta} &= k_1 - \psi \sin\phi_1 \,. \end{aligned}\right\} \qquad (15)$$

Of some interest here is a limiting value of ϕ_1 since $\phi_1 = 90^\circ$ is again a singularity.

From Equations (14) and (15),

$$\phi_1 = (q \sin\phi_0 + r \cos\phi_0) \sin\theta \,,$$

where $$\phi_0 = \int_0^t p \, dt \,.$$

Most air-launched vehicles are roll stabilized prior to firing the booster, or beginning any trajectory shaping. This diminishes the chances of flying the missile through the launch aircraft. So then q and r will be small or sinusoidal until $\phi_0 \simeq 0$. Then $\phi_1 \simeq r \sin\theta$, and so should be small unless we have high body-axis yaw rates at high pitch angles. This is an undesirable motion called "fishtailing".

The additional computations involved in this scheme have negligible effect on accuracy. The performance for the Trajectory 1 and 2 inputs are shown in Figures 12 and 13. Again, extrapolating the errors to target impact, the computer contributes about 30 and 110 feet cep, respectively.

A more interesting technique for providing all-attitude capability is gimbal shifting. Since gimbal lock always occurs only about the inner gimbal it is possible to sense the approaching problem and quickly shift the order of gimballing from yaw-roll-pitch to yaw-pitch-roll, for example. The explicit form of the transformation equations which must be solved are changed, but the complexity and accuracy are not affected. The key to this quick change is the solution of the new roll, pitch angles in terms of the old pitch, roll angles. If two equivalent column vectors of two orthogonal matrices are equal, the matrices are equal. By judicious selection of the column vectors of the old and new transformation matrices which we equate, three simple simultaneous equations are obtained. The solutions for the two new angles are

$$\left.\begin{aligned} \theta_n &= \tan^{-1}[\cos\phi_0 \tan\theta_0] \\ \phi_n &= \tan^{-1}[\tan\phi_0 \,/\, \cos\theta_n] \end{aligned}\right\} \qquad (16)$$

These equations are simple for the vector computer, and need be solved only once to shift the gimbals. If attitude output information is required for missile control, then these equations must be solved inversely and continuously at a low rate. But these computations are outside the strapped-down transformation and accuracy is not affected.

The Euler angle vector computer is operationally feasible, but is it physically realizable? - A reasonable question because the answer is "yes". The total functions of the computer were specified in detail and logic diagrams were generated. From these diagrams a preliminary hardware design study, including packaging, was completed. The functions of the computer for a typical missile are as follows:

(a) Strapped-down to inertial coordinate conversion: Euler angle, three-parameter transformation.

(b) Navigation: tangent plane.

(c) Guidance: roll attitude control, cross product pitch/yaw control in body-axis coordinates.

(d) Missile control: subsystem discretes and auto-pilot.

Although a particular navigation coordinate system and guidance scheme were selected, the choice is not restrictive. Their only purpose is to establish approximate memory size and computer solution times. Equivalent equation sets may be substituted readily.

The general characteristics of the computer are:

Type: General purpose, parallel, vector.

Number System: Binary.

Negative Number Representation: Two's complement.

Word Length: 20 bits.

Radix Point: Semi-fixed, fractional numbers.

Command Repertoire: 16 basic instructions.

Logic Type: TTL.

Memory: 512 words, 20 bit, NDRO (core rope)
64 words, 20 bit, DRO (ferrite coincident current).

Operations Rates: Add time = 2 μ sec.
Multiply = 8 μ sec.
Divide = 16 μ sec.
$\left.\begin{array}{l}\text{Rotation}\\\text{Resolution}\end{array}\right\} = 36\ \mu$ sec.

The solution time sequence of the computer is shown in Figure 14. As shown, the strapped-down conversions occur every millisecond, and the command functions are generated every six milliseconds. This is more than adequate frequency response, and, together with the 40% spare memory locations, provides good growth potential for operation with other guidance equipment. With existing logic components this computer can be packaged in 300 cubic inches and will weigh 13.5 pounds.

Although the study is concerned with a missile, the conclusions apply generally to many vehicles for which strapped-down inertial navigation is applicable. These conclusions are as follows:

1. *Cost:* Given a parallel, vector, whole-number digital computer in the vehicle for guidance and control, the strapped-down conversion may be implemented with a minor increase in memory size. Approximately 6 temporary and 12 permanent locations must be added. The exact number and length of the words depend on accuracy requirements and attitude limits of the vehicle.

In contrast, T.F.Wiener[1] estimates that the incremental computer requires 27 integrators to generate the direction cosines and perform the velocity transformations.

2. *Accuracy:* This study showed that the accuracy for this application is primarily a function of quantization which is, in part, determined by the vehicle dynamics. That the three-parameter Euler angle technique may be made equivalent in accuracy to the direction cosine methods for many applications may be inferred from the following:

   (a) A comparison of the analytic errors of the two transformations is shown in Table II.

   (b) A comparison of simulation studies for an equivalent three-axis monotonic rotation input is shown in Figure 15. The direction cosine transformation shows only truncation errors, while the Euler angle transformation, converted here to equivalent direction cosines for ease of comparison, contains all inherent errors of the method. The scatter is primarily from quantization effects. Both techniques implement only first-order integration.

3. *Flexibility:* Three items are worthy of mention.

   (a) All attitude operation is possible with no decrease in accuracy.

   (b) Attitude output information is available with no additional hardware or computations. This information is useful for the following.

      (i) Vehicle control such as roll stabilization or pitch programming.

      (ii) Vehicle attitude information for manned vehicles.

      (iii) Initialization by gimbal matching techniques. This is especially valuable for extremely short reaction time vehicles since the angles may be initialized as fast as loading a single register.

      (iv) Sensor corrections. The angles computed for body-axis to inertial conversion may also be used to convert from inertial to body-axis. For example, in initializing an inertial system to the tangent plane coordinate system, gyro fixed biases may be calibrated automatically to the limit of accelerometer sensitivity by proportional plus integral leveling loops whose output rates are converted from inertial back to gyro axes.

   (c) Integrability with position navigation systems is easier because only three parameters must be included in the state vector for the transformation process. And these parameters are Euler angles which are identical to gimbal set update quantities.

   Interfacing with the fire control computer is easier because the fire control and missile computer may both use the same number system.

For many strapped-down systems, then, a vector computer solving a three-parameter Euler transformation provides increased flexibility with comparable costs and performance.

ACKNOWLEDGEMENT

The work described in this paper was done for the Air Programs Branch of the Office of Naval Research under Contract N00014-67-C-0464.

REFERENCES

1. Wiener, T.F. *Theoretical Analysis of Gimbal-less Inertial Reference Equipment Using Delta Modulated Instruments.* Massachusetts Institute of Technology, DSc Thesis, March 1962, AD 276 254.

2. Ball, W.F. *Strapped-Down Inertial Guidance. The Coordinate Transformation Updating Problem.* Naval Ordnance Test Station, China Lake, California, NOTS TP-4267, 1967.

3. Hills, F.B. *A Study of Coordinate-Conversion Errors in Strapped-Down Navigation.* Massachusetts Institute of Technology, Report ESL-R-244, 1965.

4. Hession, R. *Analysis of a Transformation Computer used with a Gimbal-less I.M.U.* Massachusetts Institute of Technology, January 1965.

5. Wilcox, J.C. *A New Algorithm for Strapped-Down Inertial Navigation Systems.* TRW Systems, Redondo Beach, California. Institute of Electrical and Electronic Engineers, Transactions on Aerospace and Electronic Systems, Vol. AES-3, No. 5, September 1967.

6. van Koevering, A.R.
   Roth, W. *Digital Navigation Computers. Some Unique Computing Algorithms and Computer Organizations.* Institute of Electrical and Electronic Engineers, National Aerospace Electronics Conference, 17th Proceedings, May 10-12, 1965.

7. Parini, J.A. *DIVIC gives Answers to Complex Navigation Questions.* Electronics, September, 1965.

8. Hildebrand, F.B. *Introduction to Numerical Analysis.* McGraw-Hill, New York, 1956.

TABLE I

Quantization Errors

Time (msec)	Gyro Output (deg)	Computer Input		
		Pulses	Angle (deg)	Error (deg)
1	0.02	0	0.00	-0.02
2	0.04	1	0.03	-0.01
3	0.06	1	0.06	0.00
4	0.08	0	0.06	-0.02

TABLE II

Analytic Error Comparison

Error Classification	Direction Cosine	3 Parameter Euler Angle
1. Truncation (first order integration)		
Rotation Input: Drift	$\overline{\Delta\theta}^3/3$ (Ref. 5)	$\overline{\Delta\theta}^3/3$
Skew	$\overline{\Delta\theta}^2$ (Ref. 5)	0 (Note 1)
Scale	$\overline{\Delta\theta}^2$ (Ref. 5)	0 (Note 1)
Vibration Input: Limit Cycle	$\overline{\Delta\theta}^2$ (Ref. 2)	0
2. Commutation: Angle/Angle	$\overline{\Delta\theta}^2$ (Ref. 1)	$\overline{\Delta\theta}^2$
Vel./Angle	Variable (Ref. 3)	$\Delta\theta\ \dot{\overline{\Delta x}}/2$
3. Quantization	Equivalent	Equivalent
4. Round-Off	Not Evaluated	Not Evaluated

Note 1: Transformation is orthogonal

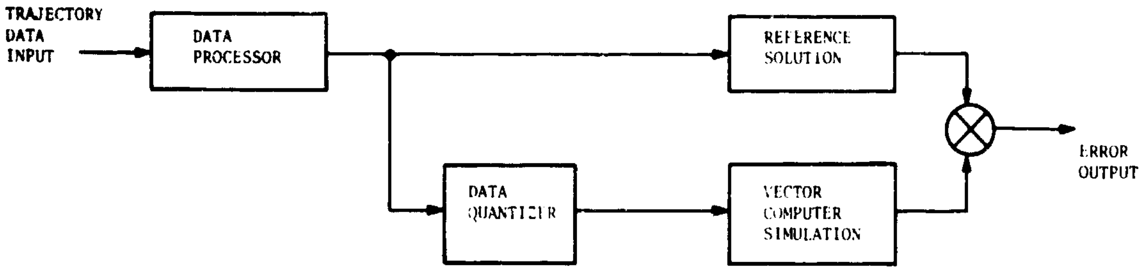


Fig.1 Simulation block diagram

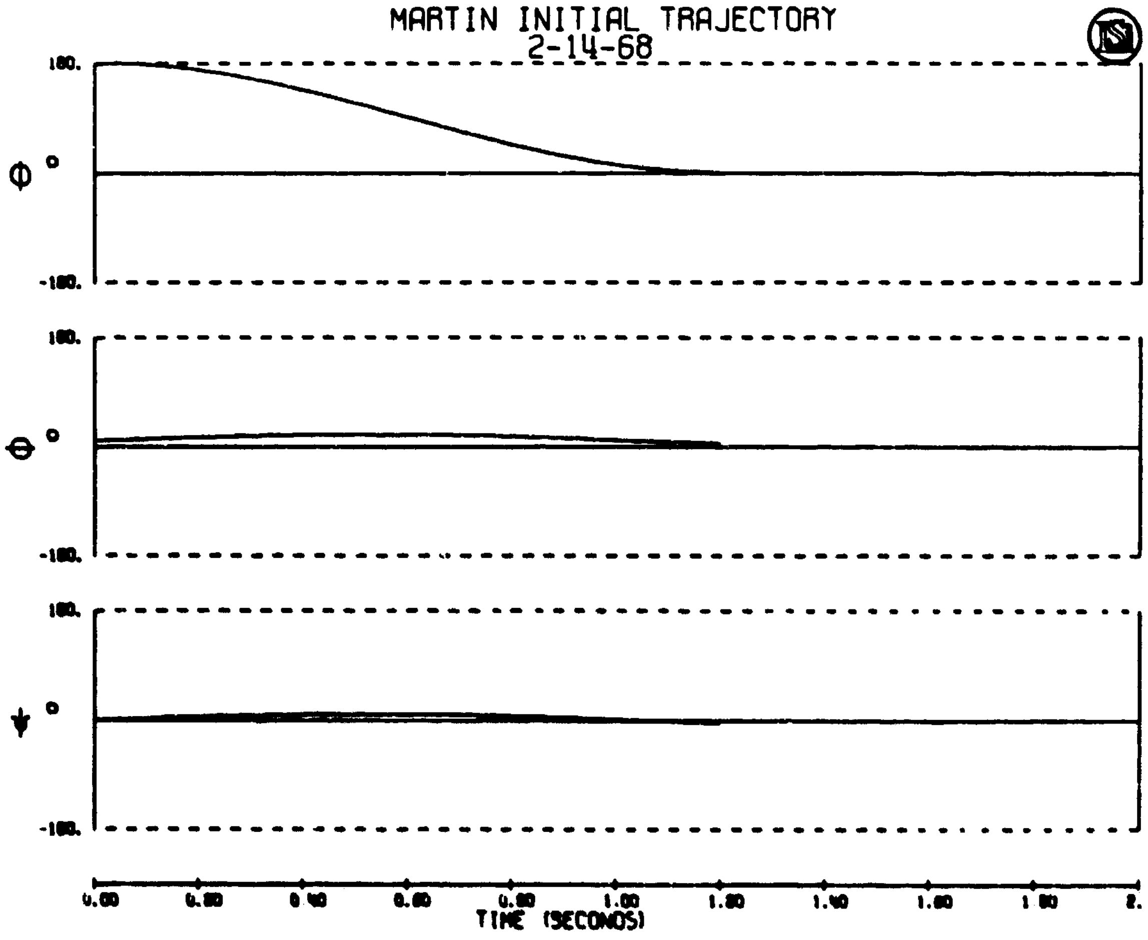


Fig.2 Initial launch trajectory (Attitude)

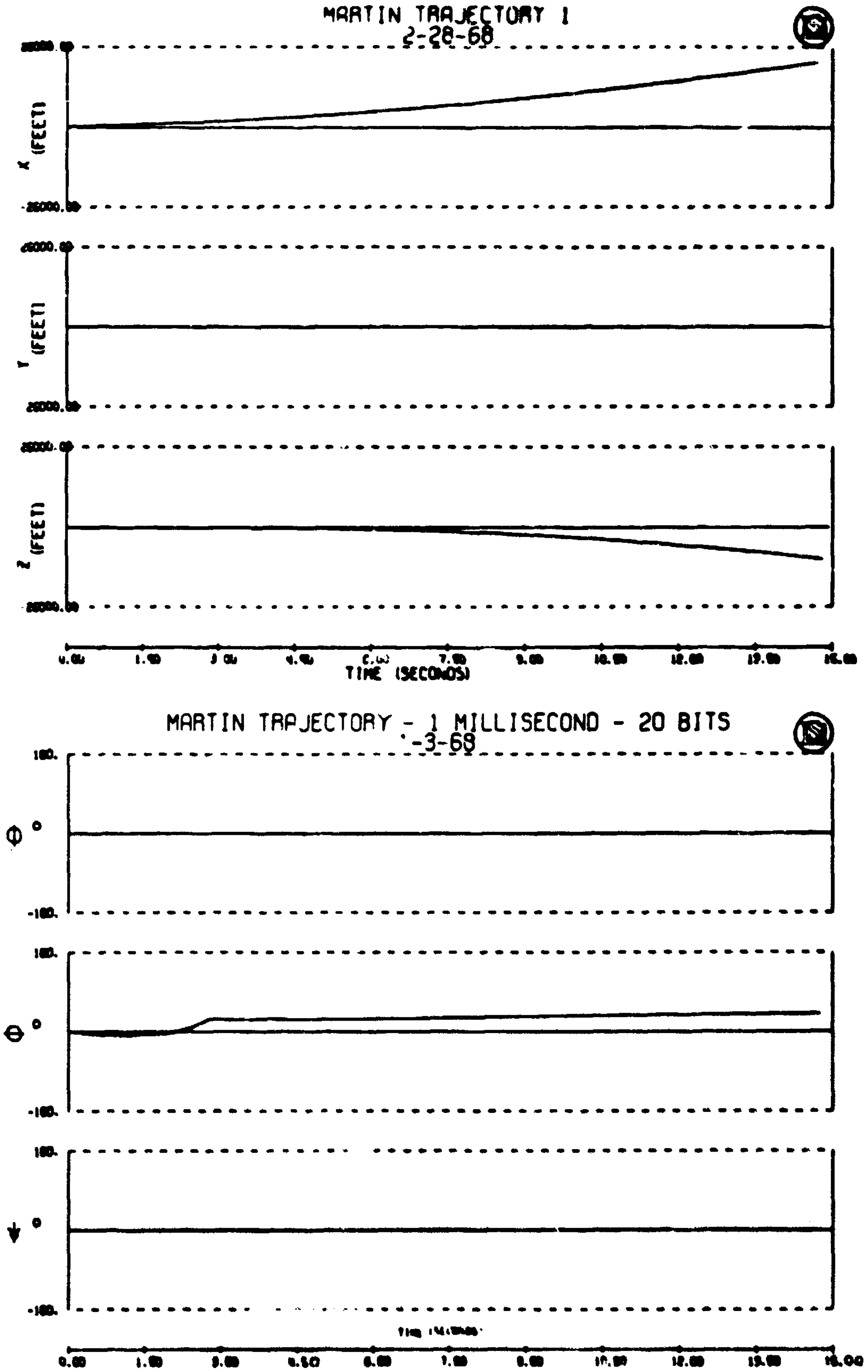


Fig.3(a) Trajectory 1 - position, attitude versus time of flight

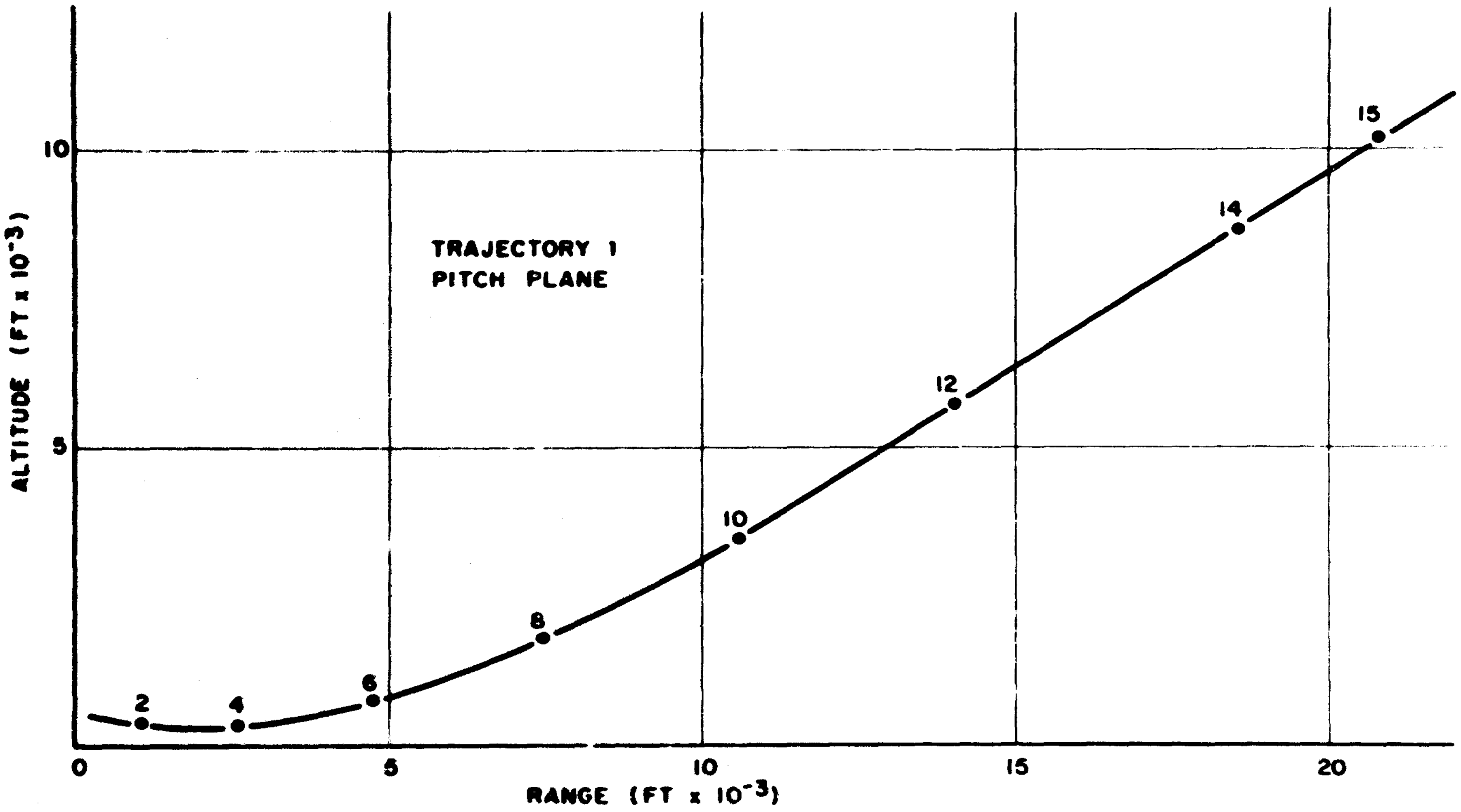
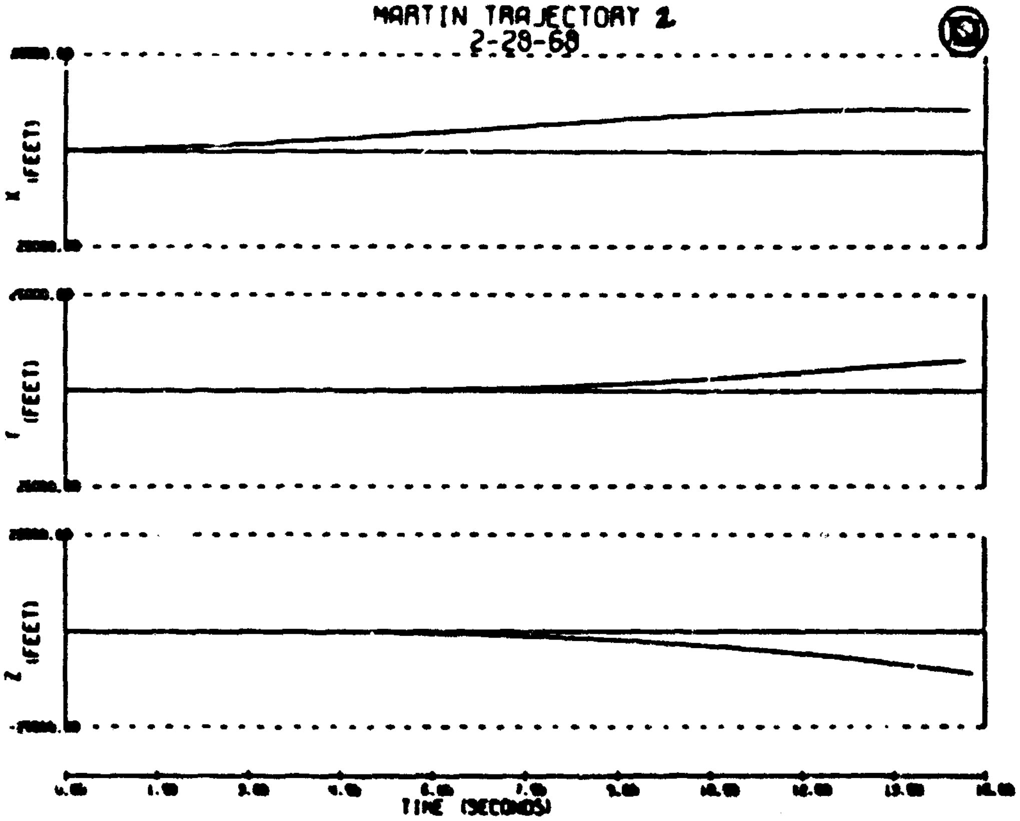


Fig.3(b) Altitude versus range as function of time-of-flight (Trajectory 1)



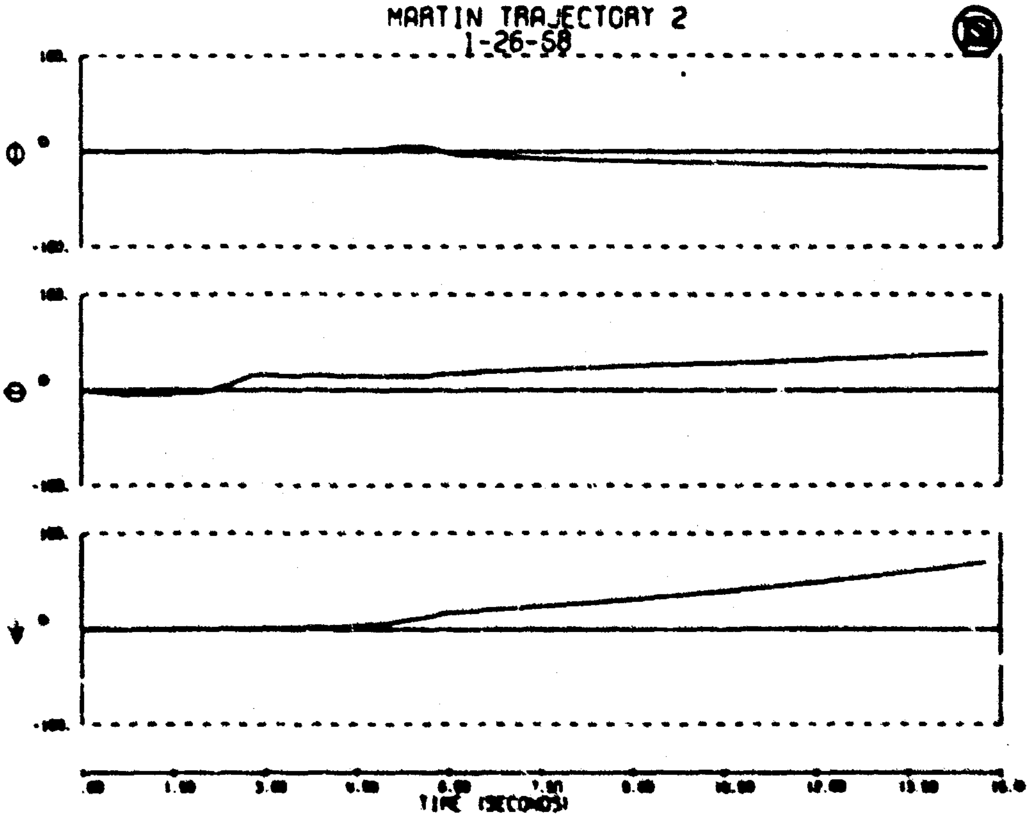


Fig.4(a) Altitude, cross range versus range as function of time-of-flight (Trajectory 2)

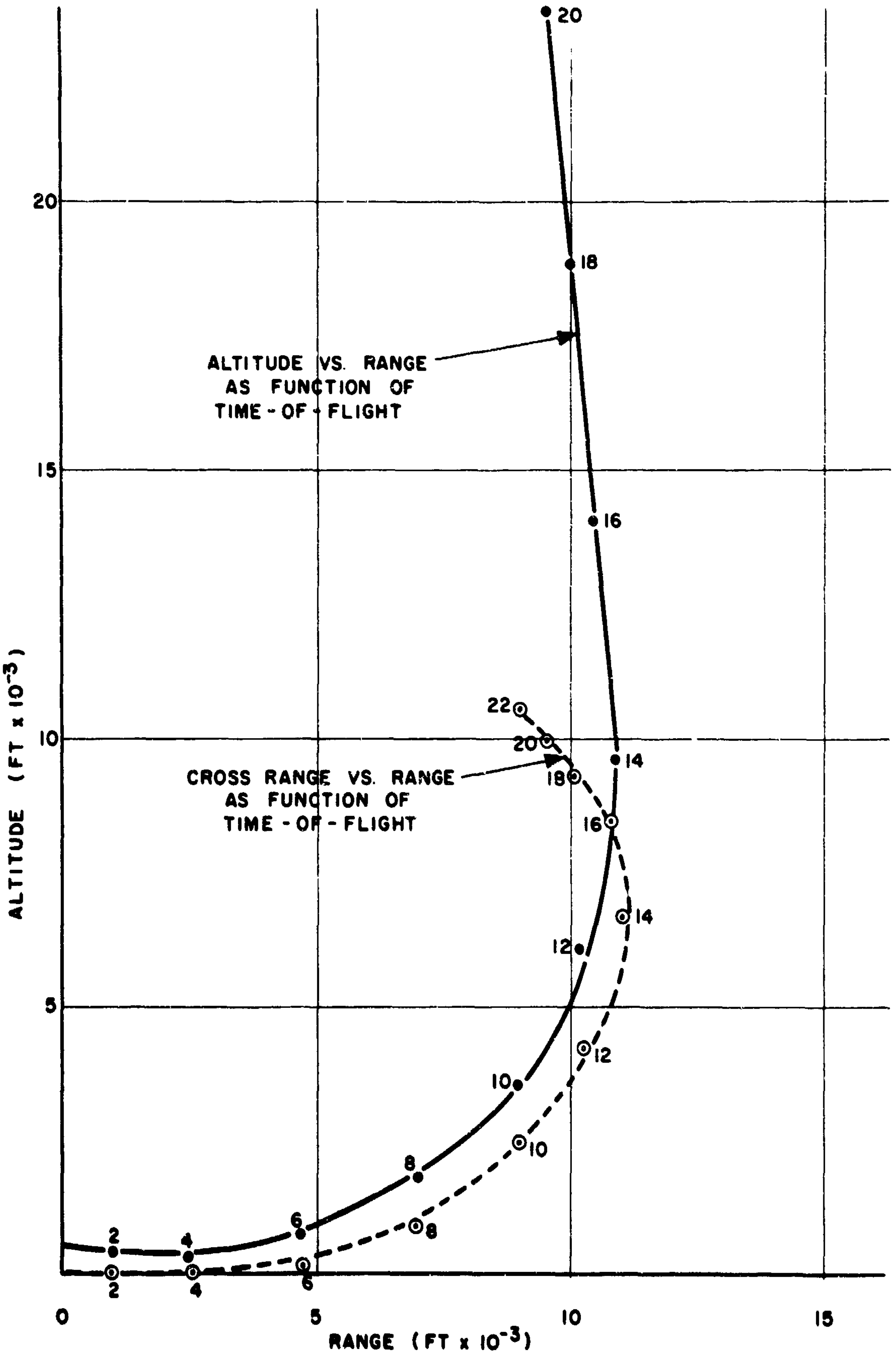


Fig.4(b) Position, attitude versus time-of-flight

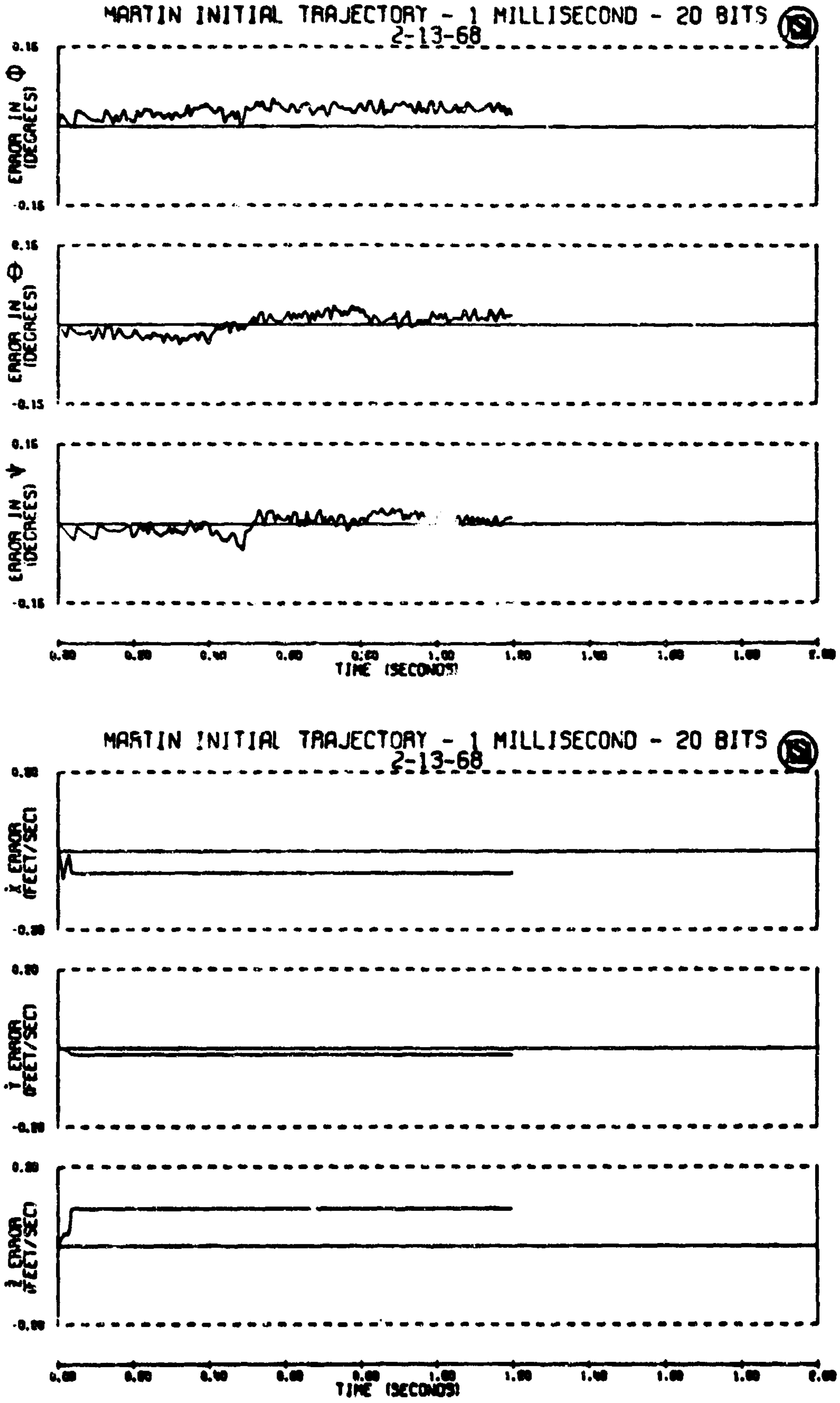


Fig. 5 Word length versus transformation accuracy

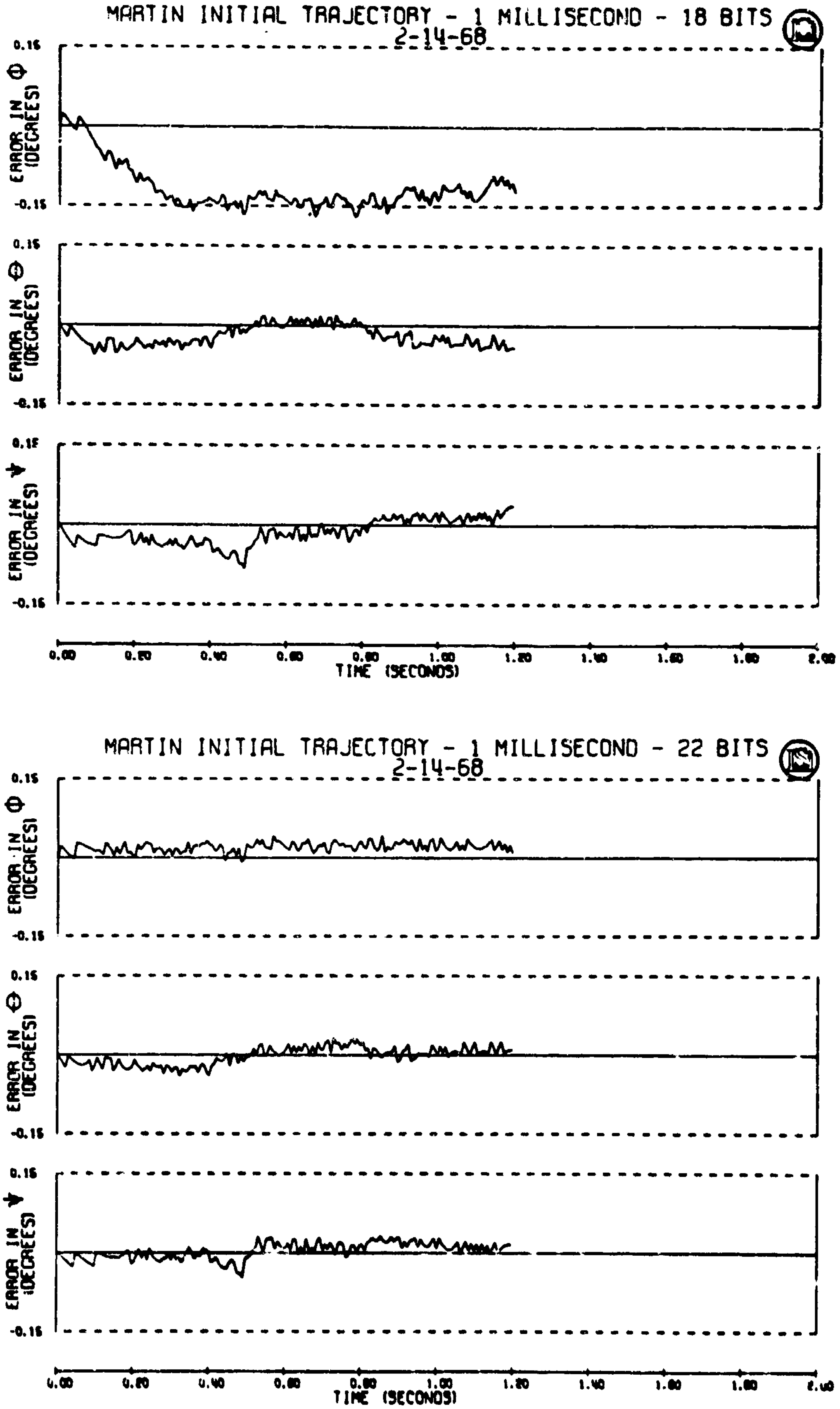


Fig. 6 Word length versus transformation accuracy

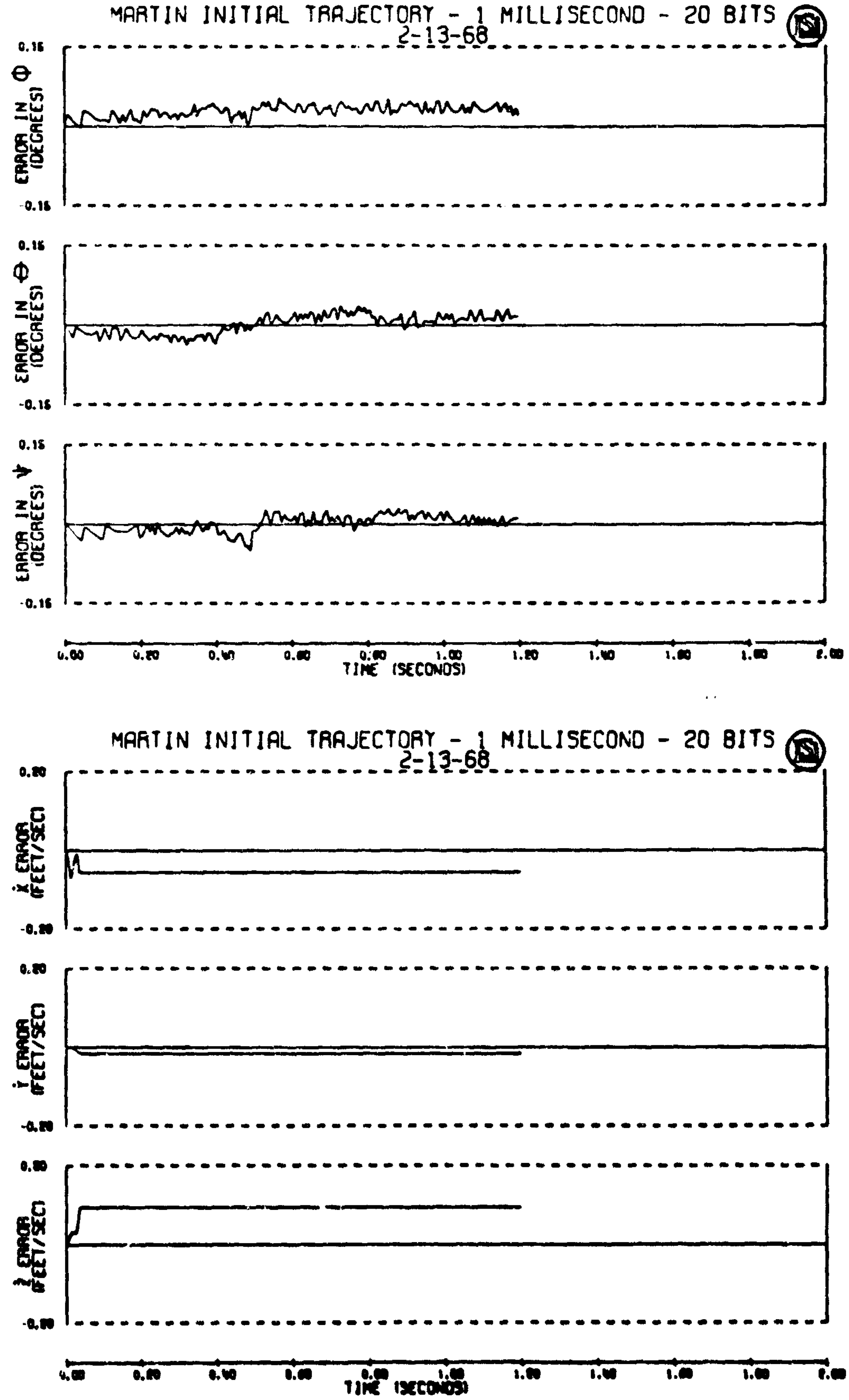
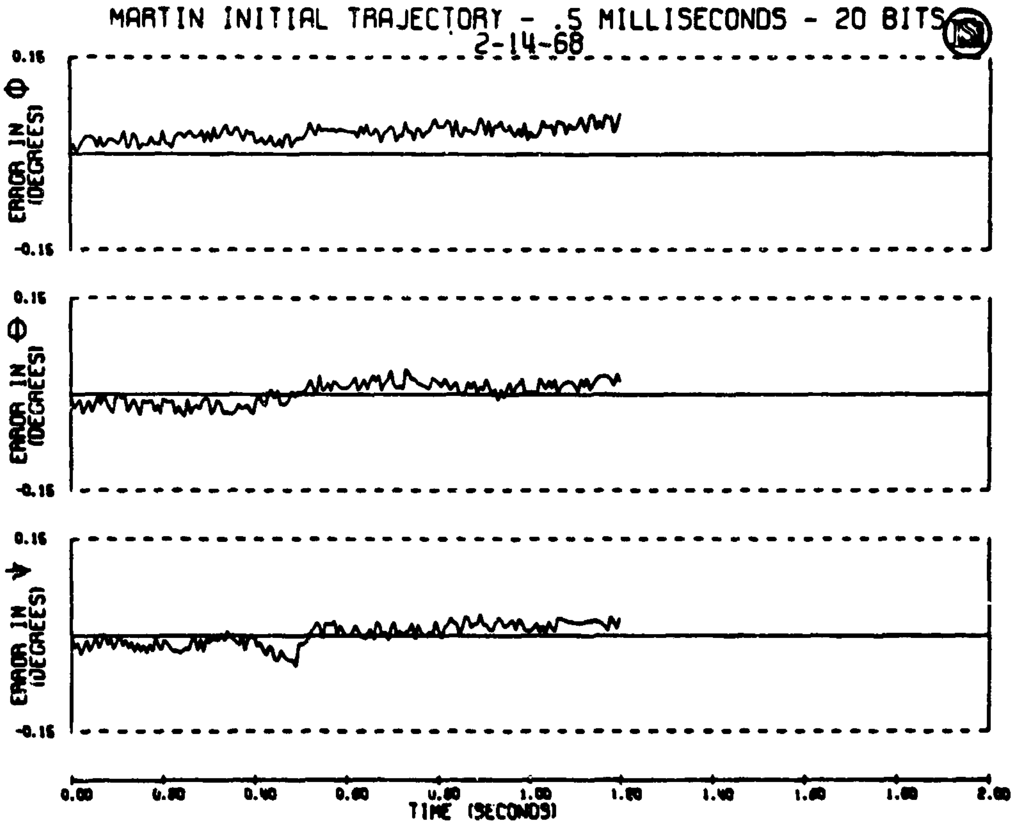


Fig.7 Solution time versus transformation accuracy



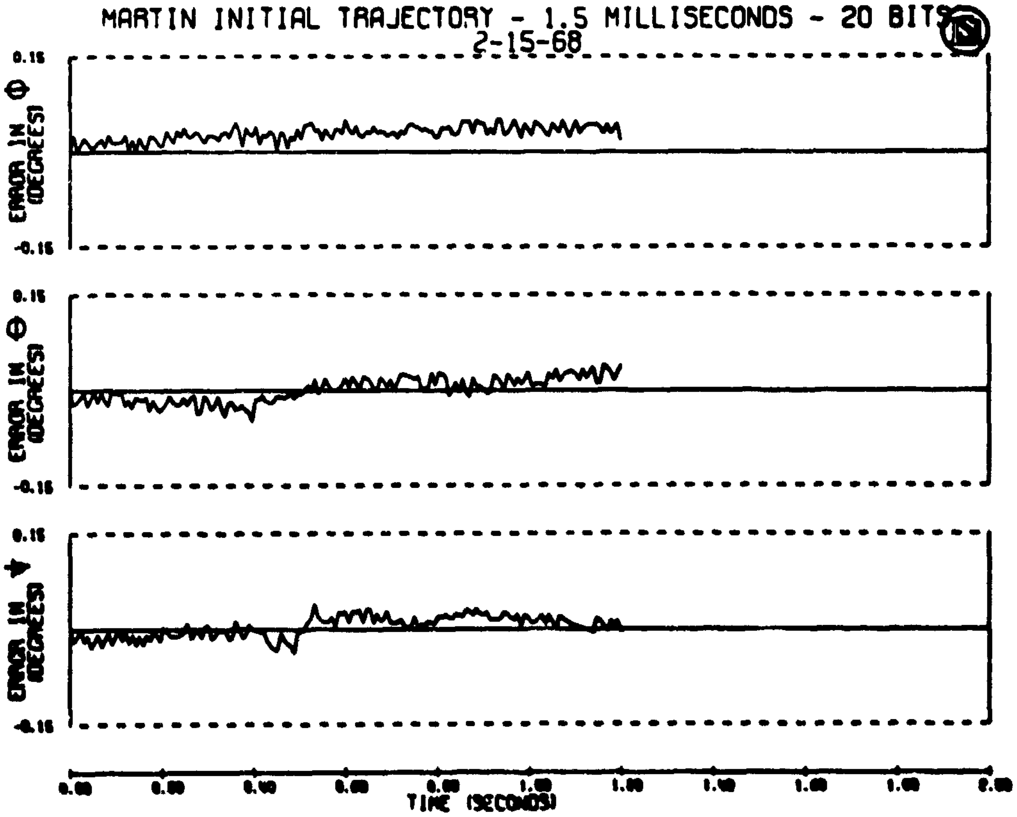


Fig.8 Solution time versus transformation accuracy

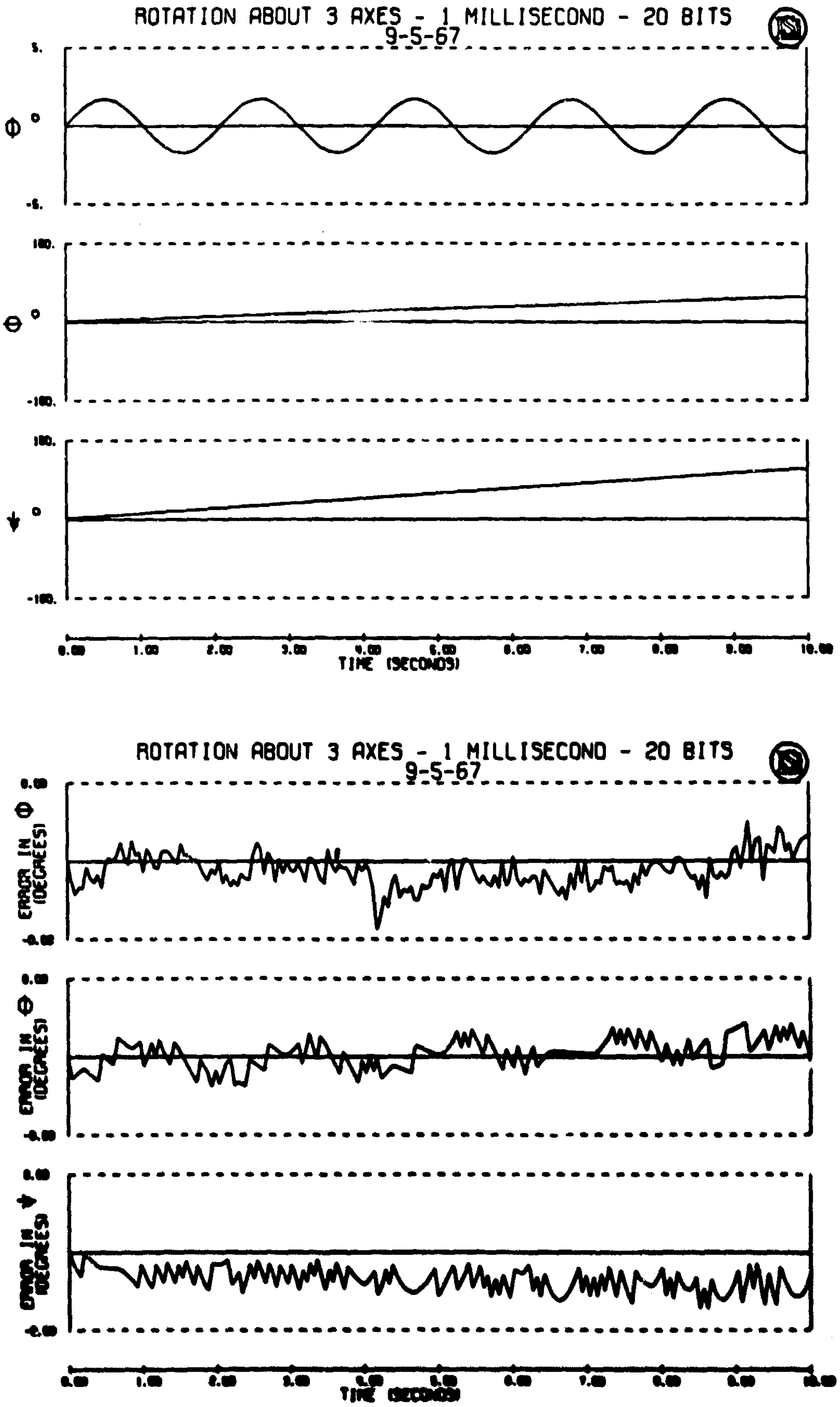


Fig.9 Complex axis rotation

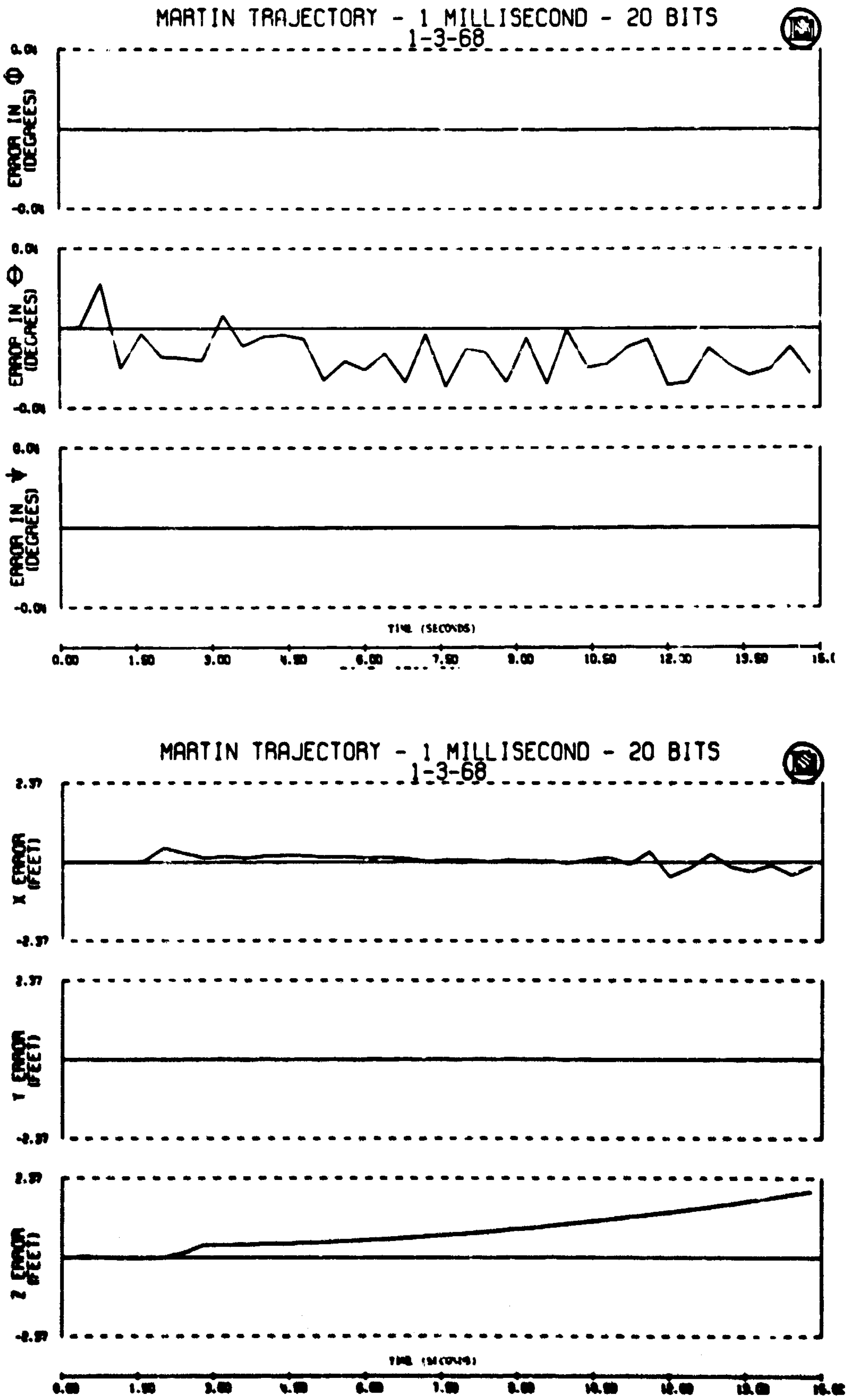


Fig. 10 Trajectory 1 - Position and angle errors, three-axis simulation

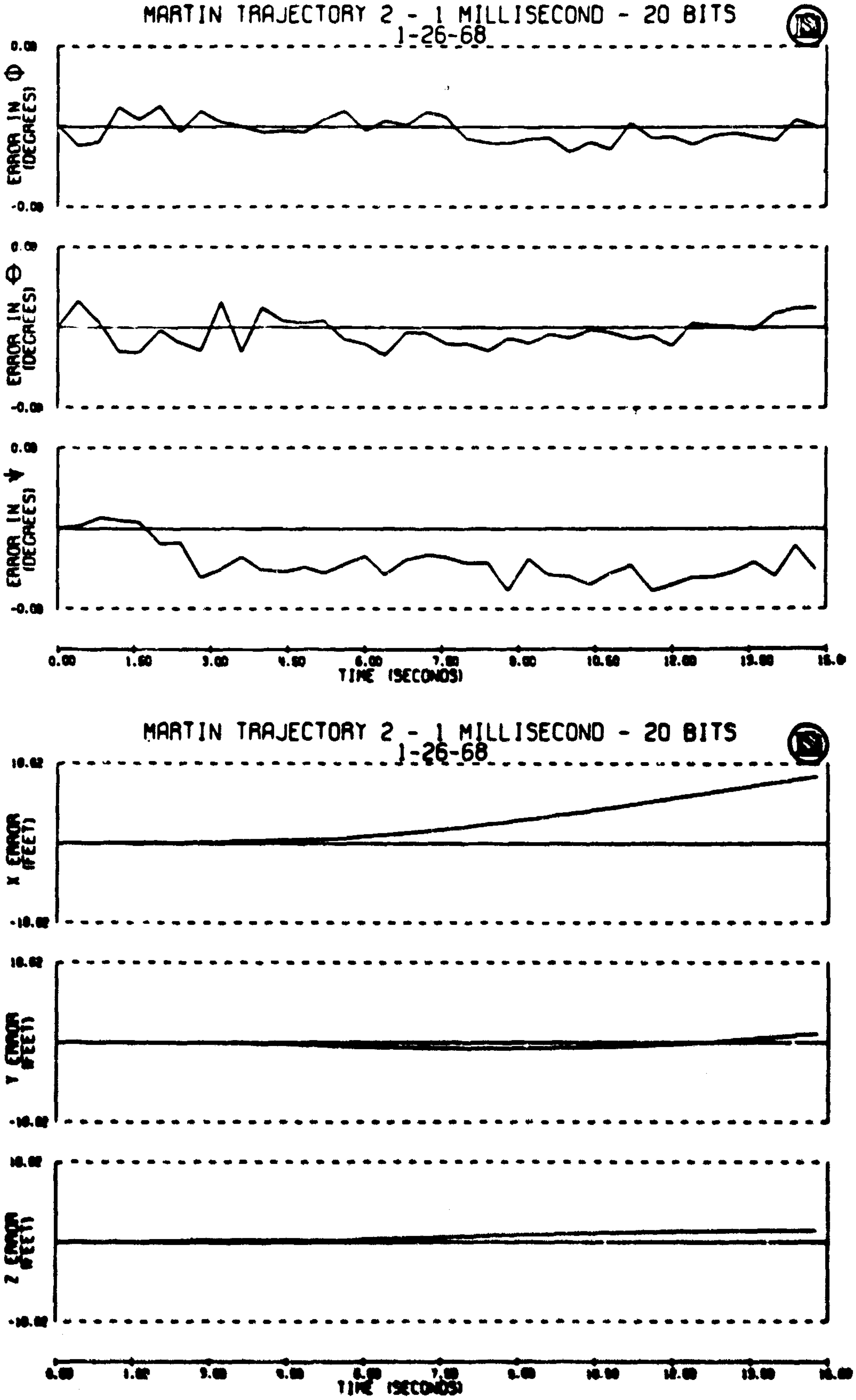


Fig.11 Trajectory 2 - Position and angle errors, three-axis simulation

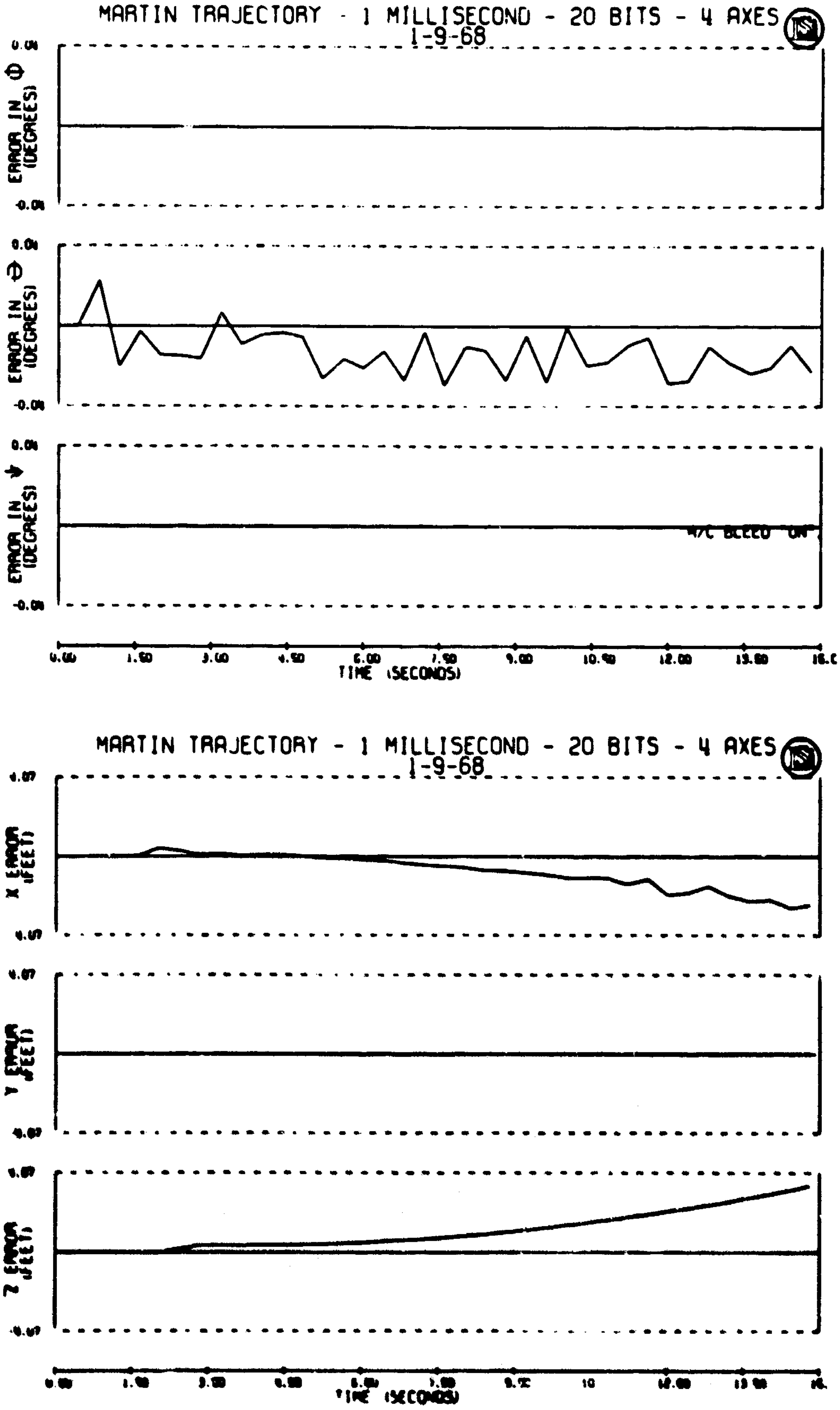


Fig.12 Trajectory 1 - Four-axis simulation

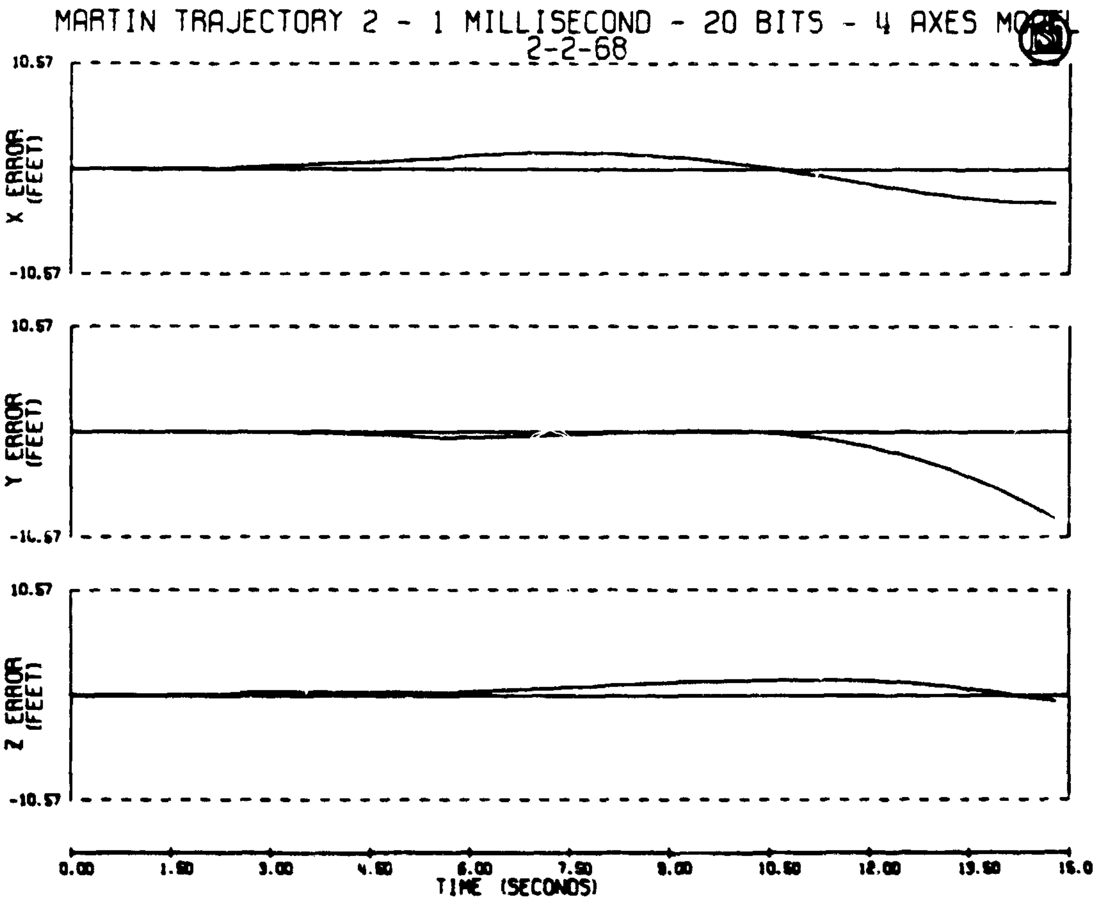


Fig.13 Trajectory 2 - Four-axis simulation

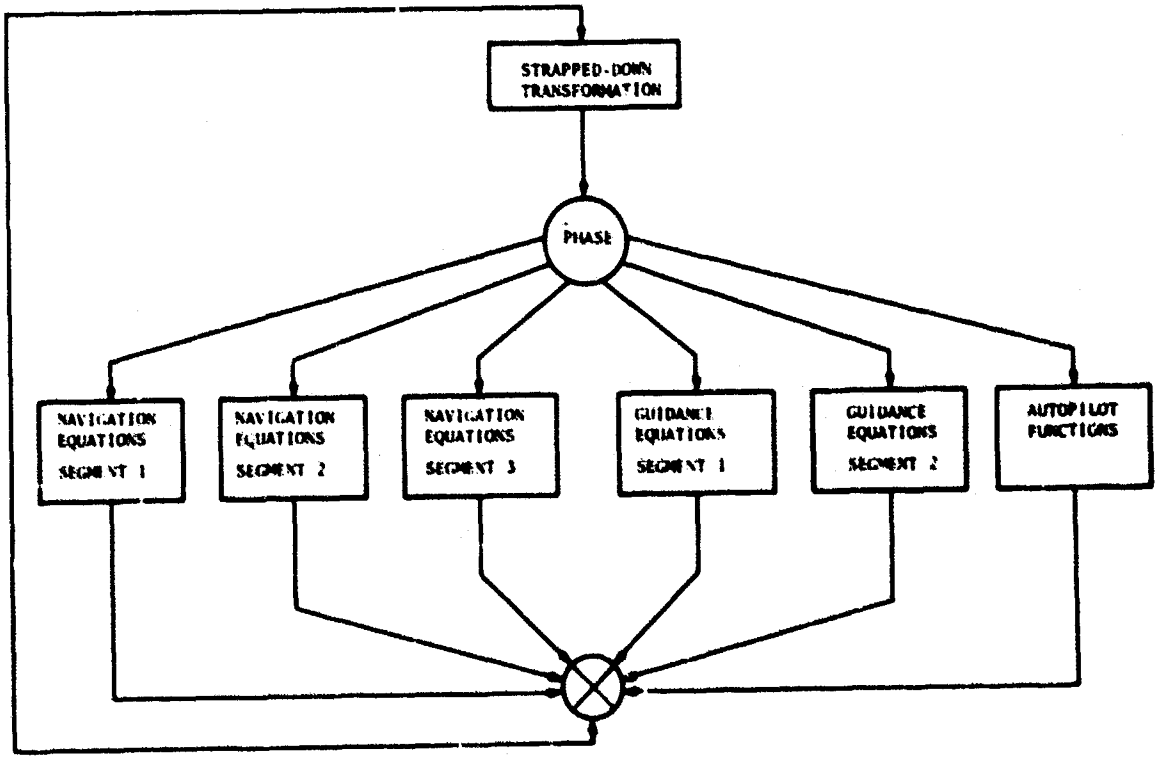
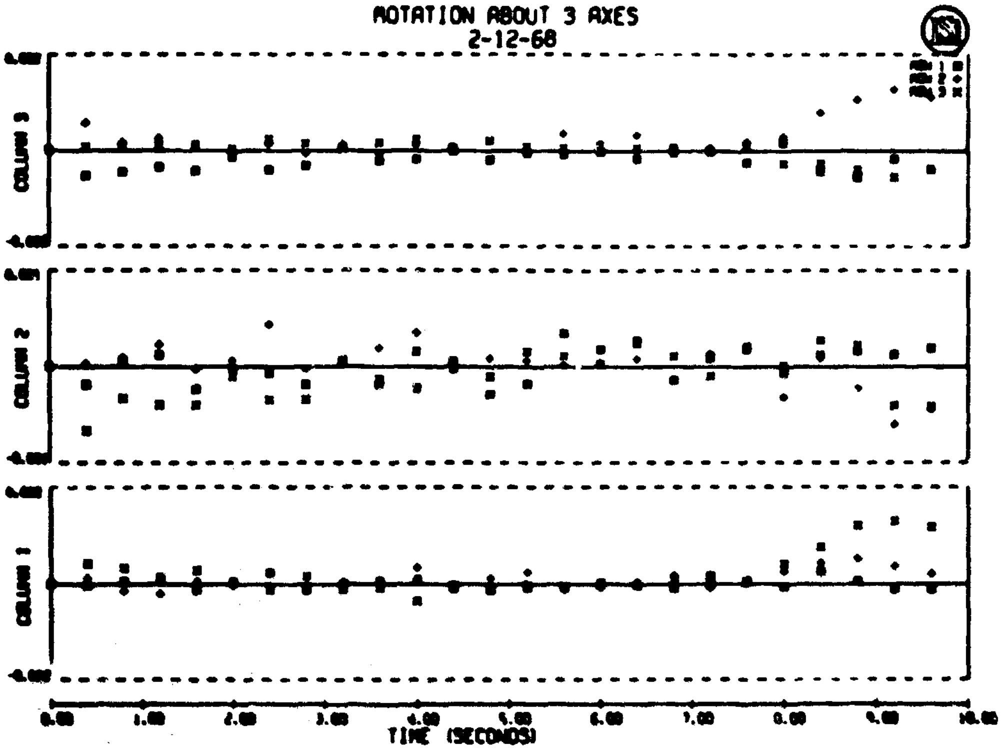

STRAPPED-DOWN TRANSFORMATION
PHASE
NAVIGATION EQUATIONS SEGMENT 1
NAVIGATION EQUATIONS SEGMENT 2
NAVIGATION EQUATIONS SEGMENT 3
GUIDANCE EQUATIONS SEGMENT 1
GUIDANCE EQUATIONS SEGMENT 2
AUTOPILOT FUNCTIONS

Fig.14 Computer time share diagram



ROTATION ABOUT 3 AXES
2-12-68

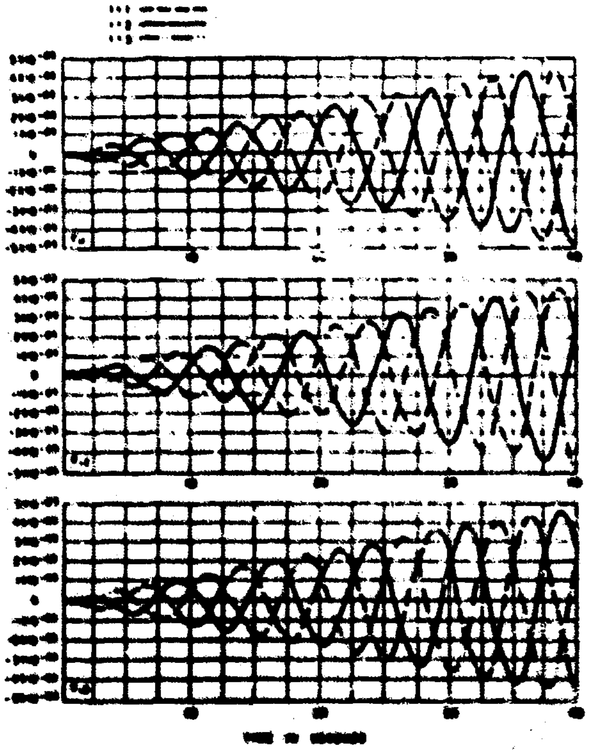


Fig.15 Euler angle/direction cosine comparison

A CONCEPT OF DIGITAL AVIONICS

by

J.F.Bussell

Elliott Flight Automation Limited,
Elstree Way, Borehamwood, Herts, England

A CONCEPT OF DIGITAL AVIONICS*

J.F. Bussell

1. INTRODUCTION

This paper is concerned with the results of studies of the application to avionics of whole-number, general purpose, digital techniques (referred to throughout as digital systems). In particular we have had the following functions in mind:

Inertial platform computations (e.g. ARINC 561).
Navigation (e.g. ARINC 562).
Automatic flight control.
Power plant control.
Head-up display.
Air data.

1.1 Simple Digital Systems

For the purpose of this paper it will help to clarify the principles of this system concept to begin by defining a basic digital system block diagram. The simplest possible system utilising whole-number, general purpose techniques is shown in Figure 1. The system consists of a number of modules :

1. *Power supply module:* This provides stable power and permits the system to be isolated from transients in the aircraft power supplies.
2. *Interface modules:* These are a wide variety of devices which are required to convert incoming data from the form in which it exists outside the digital system into the form in which it is required inside, and vice versa. These modules may have to perform analogue to digital conversions, digital to analogue conversions and/or convert from one digital format into another.
3. *Programme store module:* This is programmed for the function of the system.
4. *Data store module:* This contains the data which are normally changing during any one period of use of the system. This module may or may not be part of the programme store module.
5. *Processor module:* This acts on the data received from the interface modules, under the instructions of the programme store, producing intermediate results which are stored in the data store and putting out results through the interface modules to the displays and actuators.

1.2 The Past

Before we attempt to discuss the present and future of digital avionics it is useful to consider the past.

Early airborne digital computers were designed to meet a specific system requirement.

---

*The opinions expressed in this paper are those of the author and do not necessarily represent current company policy.

They suffered from the disadvantage that they were relatively expensive. Because of this, the idea of a so-called "central digital computer" was developed. This concept expressed the fact that the economics of general purpose computers favour their use in solving large, complex problems rather than small ones.

This idea of a central digital computer has met with considerable difficulty in aviation applications for two reasons:

Integrity

The conventional avionic system has evolved as a system of separate, largely independent units, each performing a separate function. This configuration has resulted in an inherently high integrity system whose total failure is a highly improbable event (i.e., a system which is inherently safe). The idea of a "central digital computer" challenges this integrity concept because it merges in one box the functions previously performed by many. This has the result that the probability of all systems failing is the same as for only one system. This is clearly unacceptable.

This shortcoming of the "central digital computer" concept can be rectified by introducing multiple (redundant) units in such a way as to restore the system integrity. However this must have the effect of adding considerably to

Cost

The high cost of digital systems created a barrier over which it was very difficult to sell digital devices in areas which traditionally used analogue systems. Most of the early systems were devised to meet requirements which could not be met in any other way (e.g., very high accuracy military navigation).

The basic cost factor was, and still is, aggravated by the complexity of the peripherals associated with such a system. In particular it is necessary, in most systems at present, to integrate the digital system with a wide variety of analogue transducers and actuators. This involves complicated special-to-type analogue to digital and digital to analogue converters. In the worst case the cost of these devices can exceed the cost of the other modules.

2. THE MERITS OF DIGITAL SYSTEMS

In the long term, the chief factor influencing the design of the digital systems will be their particular merits.

2.1 Flexibility

The ability to develop the system programme in a variable store results in substantial reductions in development costs. For example, in the development of an analogue autopilot, a change in the system requirement may necessitate a partial redesign of the equipment. In many cases such a change would only result in a minor programme change in a digital system. In addition to reducing the work load involved in the change, this flexibility shortens the time scale and the consequent costs of maintaining a design team.

Another important aspect of this flexibility is the ability to experiment during the development of the system without the necessity of making changes to the hardware.

2.2 Universality

The ability to offer the same basic hardware for the solution of a variety of system problems, as well as for different applications of the same type, offers a potential decrease in development costs across a wide market.

In addition to this, the large scale manufacture of such devices offers potential savings in manufacturing costs.

Finally the use of common hardware for many different functions, particularly within the same aircraft, could result in large reductions in the total cost of ownership in terms of reduced spares holdings, reduced maintenance training, shorter maintenance times and hence more efficient aircraft utilisation.

2.3 Computing Ability

Digital systems offer the ability to carry out calculations of a much higher order of complexity than analogue systems. Thus, for example, with a digital system it is possible to consider facilities for power plant control giving :

Fuel optimisation

Adaptive control

Parameter degradation and failure prediction.

These facilities can result in an effective reduction in the total cost of ownership.

One of the principal differences between analogue and digital devices is the way in which the components are used in the calculations. Analogue systems use integrating amplifiers, servos, cams, etc., each capable of performing one computing function. General purpose digital systems use most of their components, most of the time for most of the functions. This has the important advantage that the system can be programmed to carry out calculations aimed at monitoring its own performance. Such calculations can be used in two important ways :

(i) *Maintenance Checks:* Aimed at producing a rapid turn-round of the aircraft and consequently higher revenue-earning capacity.

(ii) *Failure Detection:* As part of a high integrity "failure surviving" system. Studies of this technique suggest that it is possible to achieve a very high probability of failure detection using a system which requires only a small amount of hardware in addition to the self-monitoring programme. Thus the equipment design can still be independent of the system application in the sense implied in Section 2.2.

2.4 Cost

A very significant factor in this situation is cost. A study of digital system prices reveals several striking facts.

Over the past few years there has been a very high rate of growth of digital component technology. This development is remarkable in having simultaneously produced components which are smaller, have a higher performance for a lower power and at a lower cost. At the same time the increasing complexity of each component has implied simpler system design. This factor and the increasing sale of digital systems has resulted in substantial reductions in manufacturing costs, which have also been accompanied by a rapid improvement in performance.

There is no evidence that this process of improvement in performance coupled with decreased costs has any limitations. As a result the current prices of airborne digital systems are already competitive with the most complex analogue devices.

2.5 Reliability

Because of the greater degree of standardisation, process control and automation, modern components have become much more reliable and this trend is one which can be expected to continue. The same can also be said of equipment design. Apart from these factors, the wide area of potential application offers scope for a much greater degree of reliability engineering. Thus in every respect the use of digital techniques offers considerable hope of improvement in equipment reliability.

2.6 Digital Interfaces

The use of controls and transducers with digital outputs, and displays and actuators with digital inputs would facilitate the introduction of digital data transmission systems. Such systems could be expected to result in a number of advantages:

(i) A significant reduction in the quantity and type of avionic signal wiring.

(ii) Higher integrity resulting from built-in error-detecting techniques (e.g., error-detecting codes).

(ii) The easy interchange of boxes for different aircraft roles (i.e., the location of a box can be independent of its function).

(iv) Simplified system development because of less physical interdependence of different boxes.

2.7 Conclusions

To summarise then, the present is characterised by devices whose potential in application to avionic systems is undoubted but whose development is held up by very high first cost. What then of the future?

3. THE FUTURE

Whatever is offered now by the manufacturer of digital avionics must conform to two basic rules:

(i) It must offer solutions to avionic system problems which are compatible with market requirements now. This calls for a compromise between digital technology and the commercial needs of aviation systems and expresses the practical need to produce the first mutation in the evolution of digital avionics. This requirement dictates present hardware design and implies that it must offer the cheapest solution to the smallest problems. This means that we must define the size of the smallest problem in terms of programme store capacity, data store capacity, word length, computer speed and instruction facilities.

(ii) It must offer the capability of being developed by small evolutionary steps into the ultimate system concept. This requirement recognises the built-in conservatism of aviation development which derives from the fundamental need for safety. This need is expressed in a desire to develop aviation systems from the "known" by small evolutionary steps towards the "unknown". There is a well-known word game which provides an excellent analogy to this evolutionary process. The game consists of changing one word into another in such a way that only one letter is changed at a time and that each change produces a meaningful word. For example,

WORD

CORD

CORE

CARE

RARE

RATE

RITE

BITE .

At the same time this requirement means that the design must be applicable to complex problems without seriously prejudicing the requirements of very simple systems.

Before considering the immediate future it follows from the preceding remarks that it is convenient to consider the ultimate utilisation of digital techniques. This provides an aiming mark towards which we can direct the design concept. This aspect is simplified because it need take no account of expediency but only of the ultimately compelling features of the interaction of basic avionic system design and the merits of digital systems. It is the ultimate compromise between these two technologies which we are seeking. When we have defined this ideal system we can go on to consider the evolutionary process itself and its effect on present-day design.

3.1 The Ideal Digital System

The ideal digital system would have the following characteristics:

Integrity

It would need to be capable of built-in redundancy for integrity purposes.

Digital Interface

It would make use of a digital data transmission system (see Section 2.6).

Modularity

It would be modular in the sense defined in Section 1.1, for the following reasons:

(i) The modules are designed to form part of various systems units, each of which will require additional, different, special-to-type modules and will, in general, be built into various size boxes.

While it is perfectly feasible to design the hardware modules to be compatible with standard rack sizes it is unlikely that it would be practicable to offer all systems in a standard box size.

(ii) The requirements at module level (discussed in more detail in Section 4) show that in most cases a particular module may have to have more than one form (e.g., d.c. and a.c. power modules).

(iii) High integrity requirements result in a need for multiple redundancy which is most economically employed at module level.

(iv) It is generally true that the more complex the programme the higher must be the speed of the processor. This is a situation which tends to reduce the system cost effectiveness when the required processor speeds are beyond what is readily obtainable in the current "state of the art". In this situation the only practicable solution is to adopt a multi-processor system.

(v) Modularity permits a close match between the system requirements and the hardware realisation over the broadest system spectrum.

Thus it is better to regard the modules as system tools which can be assembled together in various configurations, rather than as an unique combination assembled into a particular box size.

A typical system complex is shown in Figure 2. It should be noted that the system dispenses with the interface modules shown in Figure 1. Instead it employs a digital data transmission system, in the manner discussed in Section 2.6. The data transmission system, displays and digital modules are all arranged in multiple (redundant) configurations in each of which the level of redundancy is determined individually by the system integrity requirements.

3.2 The First Step

We must first consider what is the best way to promote the use of digital devices in hitherto analogue systems. Particular attention has been given to the functions listed in Section 1.1.

It has been assumed that

(i) A digital system must be offered as a direct replacement of an existing analogue system and must be competitive with it. Another way of stating this is to say that it is not the concern of the customer, *a priori*, that the system is analogue or digital but only that it does the right job, in the right way, at the right price. This implies that it is unwise to use the merits of digital systems to counter disadvantages in cost, size, weight, power consumption, reliability and performance.

(ii) For the previous reason, and because of the inherent need for caution in system design, it is unwise to modify system integrity concepts until experience has been gained by substituting digital devices in systems where the integrity concept has been well proven with analogue devices.

Thus we conclude that the first step is to specify minimal modules which are competitive with existing analogue devices.

4. THE SYSTEM CONCEPT

Arising out of the various considerations already discussed, we have developed a system concept. This may be simply defined as a digital, avionic, system concept designed to match the system costs as closely as possible to the spectrum of market applications and aimed at providing the first step of an evolutionary development towards the ultimate, digital, avionic system.

To provide for the simplest system applications and, in particular, to promote the first applications of this type of digital system it is necessary to optimise the design around the simplest system requirements. This, together with the reasons enumerated in Section 3.1., leads to the conclusion that the concept must be modular. For the simplest systems, minimal cost is achieved by providing for a small store module and a simple processor. The more complex requirements can be met by multi-store, multi-processor systems. From the point of view of the processor it can be stated that minimal cost hardware necessitates optimum programme facilities. That is to say that we cannot achieve a minimal cost solution without a detailed study of the functional requirements of each calculation. Such studies have been carried out and have allowed us to arrive at a statement of the functional specification of the various modules. These are now discussed.

4.1 Power Modules

For aviation requirements it is likely to be necessary to provide miniature stabilised power supplies operating from a.c. or d.c. aircraft supplies. In addition non-miniature modules operating off mains supplies would be required for ground rigs.

4.2 Interface Modules

The system concept implies the ability to interface with many different systems and to be capable of fast data rates over a limited period. The module interfacing system must be capable of accepting special system oriented interfaces (e.g. analogue-digital-analogue converters) and to have direct access to the data store, or to be controlled by the processor.

4.3 Store Modules

One of the commonly accepted features of digital systems is the great flexibility offered by programme stores in which the programme is written electrically (see Section 2.1). However, this flexibility presents a risk of failure in the presence of electrical disturbance. In view of the high integrity requirements implicit in many of the functions considered, it is regarded as essential to provide fixed (i.e., mechanically determined) programme store modules. This requirement is facilitated by the fact that these applications are characterised by programmes which require small amounts of work space in relation to the programme storage and which are likely to remain fixed throughout the useful life of the system.

In practice, however, the situation is less clear cut and there are numerous conflicting requirements. For example:

(i) The advantages of variable store for development purposes, already mentioned.

(ii) There are many systems where data is semi-permanent, in the sense that it may need to be changed infrequently during the life of the system.

(iii) In some systems there may be a requirement to modify the programme during the working life of the aircraft.

(iv) There may be a requirement for different programmes for different roles (e.g., in military aircraft).

In practice then it would seem that we must accept that integrity varies with the degree of permanence of the programme and that all systems will, in general, require a mixture of storage media of varying permanence in differing proportions. Clearly it would be desirable to define a storage system which would offer all types of storage within the same basic module. This would have two principal advantages. Firstly there would be complete interchangeability between development and production versions of the same system. Secondly it would minimise the hardware required for store driver circuitry.

On the basis of various detailed studies of the applications listed in Section 1.1, it has been concluded that the optimum programme store module size (consistent with the processor facilities described below) is 4096 words. A smaller store is likely to restrict one or more of the applications considered.

These studies have also shown that, for this programme store module size, the minimum data store is 128 words.

4.4 Processor Module

The principal factors affecting the quantity of hardware in the processor and the data store are the instruction facilities and the word length(s) to be used.

4.4.1 *Instruction Set*

The number of instructions which it is possible to mechanise in a digital processor is a function of the number of registers in the processor. Even for a minimal type of processor, such as considered in this paper, it would probably be possible to mechanise something like 30 to 40 different instructions. It follows that the number of ways in which these instructions may be combined to give an instruction set is almost infinite and it is clearly impractical to study and compare all possibilities.

The approach which was therefore adopted was to study the utilisation of orders in other computers of a similar specification in an attempt to determine which were really effective. The effect of removing orders singly was studied in terms of the effect on programme length. The removal of one order and the substitution of another was also studied in terms of its effect on overall programme length and cycle time. By this method it was possible to justify the inclusion of each order in terms of its effect on programme length and/or speed.

Studies of the instruction set required have revealed a number of significant facts about the utilisation and efficiency of various types of instruction. Figure 3 shows histograms illustrating the distribution of instructions in typical military navigation programmes using the Autonetics Verdan computer and the Elliott M.C.S. 920 computer. The important conclusion to be withdrawn from these two histograms is that the utilisation of different instructions is very variable. This suggests that, beyond a certain basic minimum number, the addition of further instructions to the basic set is unlikely to be justified on grounds of utilisation alone. These studies suggest the minimum instruction set of Table I. (In the table the effect on programme capacity of removing orders singly from the instruction set is referred to as the "effective capacity value".)

TABLE I

Order	*Reason for inclusion*	
WRITE JUMP IF NEGATIVE NEGATE & ADD INPUT/OUTPUT	Essential basic orders	
READ	Effective capacity value (%)	18%
ADD		10%
ORDER MODIFY		8%
JUMP (Unconditional)		7%
SHIFT		7%
SUBROUTINE ENTRY		5%
EXCHANGE		3%
JUMP IF ZERO		2½%
MULTIPLY	Omission reduces the effective speed of the processor by a factor of	5
DIVIDE		4
STORE EXTENSION REGISTER	Facilitates extended length working (see Section 4.4.2)	
COLLATE	Facilitates efficient use of data store	

4.4.2 *Word Length*

The data word length determines the accuracy with which the processor can readily perform any calculation. The word length is defined as the number of binary digits or bits. If a word length of n bits is used, a number can be represented to an accuracy of one part in 2^n. It is a convenient rule of thumb that 10 bits corresponds to 3 decimal digits (i.e., $2^{10} = 1024 \simeq 10^3$).

If the processor is not to degrade the accuracy of the system the word length must be consistent with the accuracy of the transducer data entering the system. In general, in avionics, the data is seldom as accurate as 0.1% and so 10 bits is adequate. Allowing a further 2 bits for round-off errors in calculation gives a satisfactory working word length of 12 bits.

This is not an absolute limit, however. There is nothing to stop the processor working to much greater accuracy by using special programmes. To facilitate this extended length working it is necessary to provide the accumulator with an extension register which can hold the least significant half of the result of a multiplication. The contents of this extension register can be written into the data store using the "store extension register" instruction. The register can be loaded as a "side effect" of another instruction (e.g., negate and add).

4.4.3 *Instruction Structure*

With a system utilising a relatively small data store it is possible to use a fixed programme store more efficiently by adopting a two-format instruction structure in which short words are used for instructions addressing the data store and long words for programme jumps and the storage of constants. Such a system can result in a 10% improvement in programme store utilisation. However, in the more general situation, where some systems require large amounts of variable storage, the scheme is inefficient. For this reason a single format structure is adopted.

Further than this, for systems using large amounts of variable store it is desirable that data and instruction words are interchangeable. For this reason it is concluded that the instruction word length should also be 12 bits.

To obtain the instruction set in Section 4.4.1 efficiently it is necessary to allocate 4 bits to define the instruction function. This leaves 8 bits only to define addresses. Addressing the whole programme store module can be arranged by the use of "order modify" facilities. Address extension registers built into the additional store modules can accommodate multi-store systems.

4.4.4 *Speed*

Assuming we adopt the same techniques, a more complex calculation will require a longer programme and will take longer to perform in the computer.

In general it is possible to increase the speed of a programme by using more programme store. For example a large stored table is often a faster method of generating a function than a small polynomial subroutine which occupies much less programme store capacity. Thus larger storage systems can be utilised to increase the effective speed of the system as well as to increase the size of the task performed.

Speed requirements are basically determined by the function which the system is required to perform. For the system functions listed in Section 1.1, it has been found that in the worst case an average instruction time of about 7.5 microseconds must be achieved. (Note that this average is taken from the programme and not from the instruction set.)

Another factor contributing to the effective speed efficiency of the system is the interrupt facility. Many of the system tasks envisaged are either time-dependent, in that a defined and fairly exact repetition rate is required, or alarm-dependent, in that the programme must react quickly to a relatively infrequent high priority stimulus. These requirements indicate the need for an interrupt system, such that the base programme may be interrupted at any time and a more important task may be undertaken. The hardware interrupt system must operate either with a simple hardware interrupt discriminator or with an executive programme which will assign task priority to the interrupting device. As the interrupted programme must always start at a given point and cannot itself be interrupted, the provision of a fixed interrupt level start address is acceptable. This removes the need to store the interrupt level sequence control register and reduces the processor hardware. The interrupt discrimination can be provided by allowing the fixed interrupt level starting address to be externally modified. The logic required to operate this discrimination system can be system oriented and can form part of the system oriented interface logic.

5. CONCLUSIONS

The purpose of this paper has been to describe the basis of the development of an evolutionary, generic concept of minimal cost digital avionic modules. The concept is one which extends across the whole range of avionic computation and can also be regarded as part of a wider concept including industrial process control systems.

There is no doubt that the potential merits of digital systems offer the prospect of an ultimate revolution in avionics. The system concepts eventually adopted will look very similar to those conceived for ground applications. However, revolution is unacceptable in avionics and these changes cannot be expected to occur quickly. For this reason the concept discussed in this paper is one which is essentially evolutionary in character. It offers both the long-term system advantages and the essential first step on the road towards ideal digital systems.

It is only within the last few years that digital devices have reached the point where they are likely to be economically viable, and then only with the more complex analogue devices. It is probably fair to talk in terms of digital computer price reductions of 80% within a decade. These reductions result in the main from very rapid changes in component technology. These changes have, in most cases, simultaneously offered reductions in size, weight and power consumption, as well as offering improvements in functional performance and reliability, a situation which must be almost unique in technological history. What is more, this process shows no sign of slowing down. For this reason it is possible to predict with confidence a continuing reduction in the real price of digital devices. The rate of this reduction is likely to outstrip any corresponding reduction in the price of analogue devices and so it is reasonable to anticipate that digital devices will be viable in a larger and larger sector of the market.

In addition to the factors just mentioned, the increasing use of digital devices in aircraft is likely to promote the development of transducers with digital outputs and digitally driven actuators and displays. This will create an environment in which digital data transmission and display systems are encouraged to develop. This in turn will result in a dramatic reduction in the need for analogue to digital and digital to analogue conversion modules which currently represent major cost and reliability obstacles to the sale of digital systems. As a result it is possible to predict an acceleration of the utilisation of digital devices, moving progressively towards the ultimate concept of multi-processor, multi-store systems associated with digital data transmission and display system complexes.

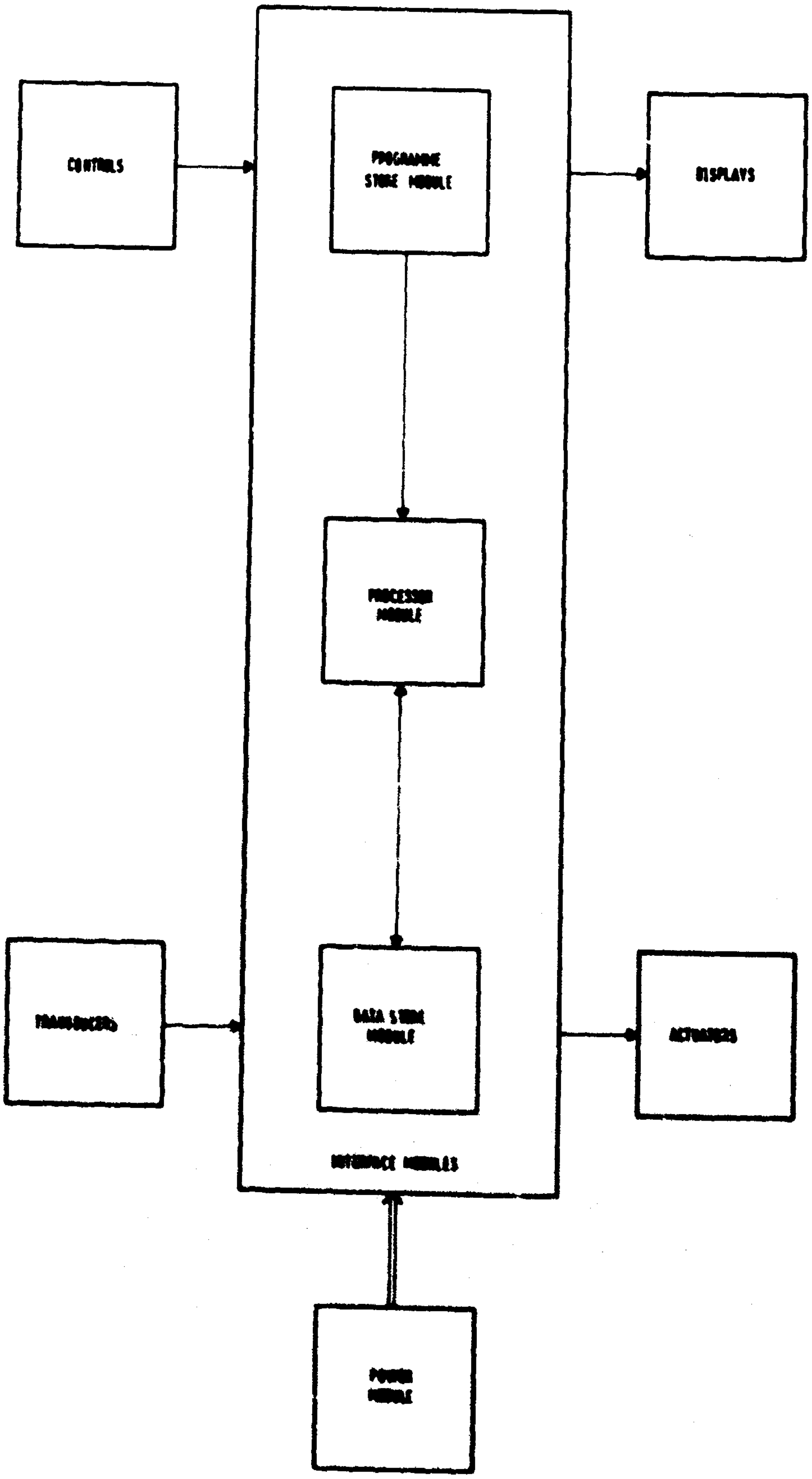


Fig.1 Block diagram of a simple digital system

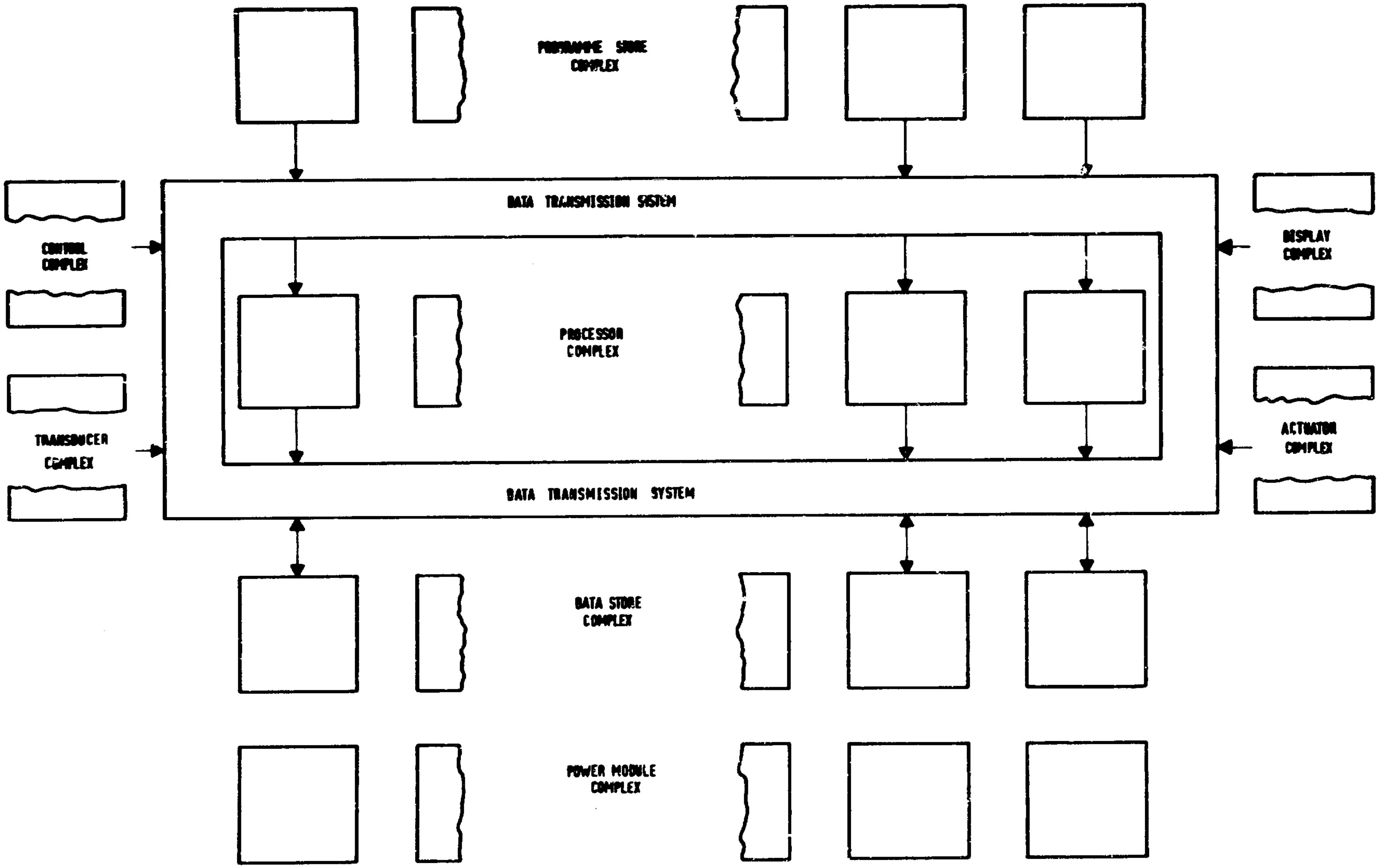


Fig. 2 Block diagram of a digital system complex

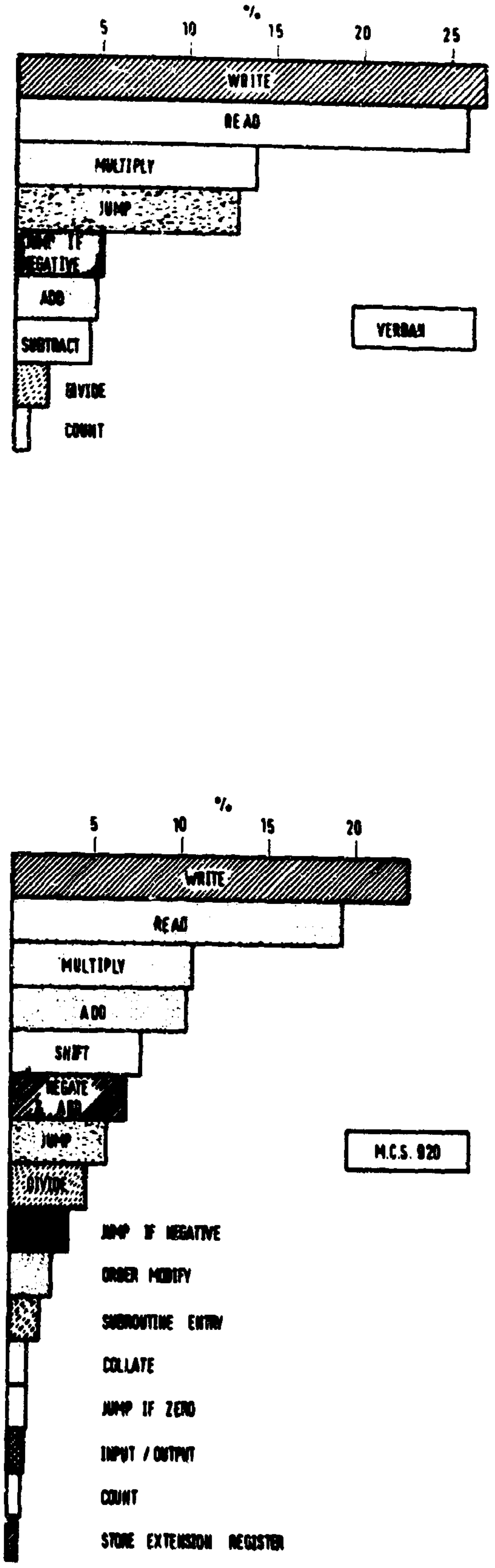


Fig.3 Utilisation of programme instructions

MICROPLASTICITY IN METALS FOR PRECISION INSTRUMENTS

by

O.H.Wyatt

Royal Aircraft Establishment,
Farnborough, Hampshire, England

SUMMARY

The microplasticity of metals is discussed and work on materials used in precision instruments is reviewed. Some new data are presented for beryllium obtained from R.O.F. Cardiff, which closely agree with results on beryllium from the Brush Beryllium Corp. Variable drift in a precision gyroscope has been traced to ligament creep and steps to reduce the effect are described.

MICROPLASTICITY IN METALS FOR PRECISION INSTRUMENTS*

O.H.Wyatt

1. INTRODUCTION

Components for inertial navigation systems can be regarded as very precise arrangements of metals, ceramics, glasses and polymers. A great deal of attention is paid to optimising the arrangement, considerably less to the properties of the materials. This may be due to the neglect of materials science in most engineering and physics curricula and the noticeable unwillingness of metallurgists (or materials scientists, as they are now known) to take engineering-type jobs, although their knowledge and abilities would be an important contribution to many engineering projects. A count of the number of different materials in the sensitive element of one gyroscope gave over 20 metals, 7 polymers, 1 cermet and 1 glass. Each of these possessed properties vital to its application, but detailed questions on their structure and properties often go unanswered for lack of trained opinion.

In planning this paper it was intended to cover briefly all classes of materials and show how the different types of atomic bonding are responsible for their macroscopic properties which make them suitable for particular applications. This turned out to require a text-book, not a paper, and the field has had to be narrowed to metals, and in particular, their microplasticity, and anelasticity, that is, small departure from Hookean elastic behaviour (stress = constant × strain). The topic is one of three legs on which the larger subject of material stability rests, the other two being residual stresses and phase changes. It is the intention to cover these in a later paper.

Microplasticity is peculiar to metals as a class on account of the ease with which dislocations can move in a crystal or grain. As is well known to materials scientists, dislocations are the line faults in the regular arrangement of atoms in a crystal which are responsible for the slipping of atoms planes past each other during plastic flow. Slip is inhomogeneous like the forward progress of a caterpillar. It starts on some part of an atomic plane and the dislocation separating the slipped and unslipped regions moves steadily over the plane. All materials have large numbers of dislocations built into the crystal structure (10^5 - 10^7 cm of dislocation per cm^3 of material) and under stress they begin to move. Because of the nature of ionic and covalent bendings present in ceramics, dislocations require high stresses to move them and ceramics usually fracture before they flow plastically. Metallic bonding, however, permits the dislocation to move over many of the atomic planes under low stress and the problem with metals is to prevent dislocation movement, so as to give a useful elastic range. There are a limited number of techniques available which, for simplicity, may be regarded as distorting the regularity of the lattice by introducing oversize solute atoms or fine particles of a second phase. These methods are only partially successful in holding the dislocations in position until the applied stress is sufficient to cause large-scale dislocation movement and gross plastic flow.

---

*This paper expresses the opinions of the author and does not necessarily represent the official views of the Royal Aircraft Establishment.

Mechanical designers often assume that metals are ideally elastic, as envisaged by Hooke (1678), up to the yield stress above which gross plastic flow occurs. This is adequate for many purposes but, even without any knowledge of the physical mechanisms involved, it is obviously inadequate in others. It is well known, for example, that repeated application of stress well below the conventional "yield stress" will eventually cause fatigue fracture, indicating that ideal elastic straining is not taking place. Again, under creep conditions, the strain is a function of time although this parameter does not appear in Hooke's Law Much of the work of physical metallurgists is devoted to microplasticity in one guise or another and it is only possible to select certain topics here which seem most relevant to the design of I.N. components.

The following sections discuss various forms of microplasticity: microyielding, hysteresis, damping and microcreep. Then follows a review of previous experimental work, mainly on microyielding, with materials of interest to component designers. The final two sections relate some recent work on beryllium microyield stress and on ligament microcreep.

2. RELATED EFFECTS IN MICROPLASTICITY

Several closely related effects are associated in microplasticity, as illustrated in Figures 1, 2 and 3. Figure 1 shows a conventional stress-strain curve of a ductile metal: a nominally linear and elastic strain up to about 10^{-3} is followed by gross plastic flow. Unloading at point 1 on the yield curve brings the strain down the line 12 approximately parallel to the initial elastic range. To avoid dispute as to the exact point of departure from linearity, the yield stress is defined as the position at which 0.1% strain off-set from the tangent to the initial part of the stress-total strain curve has occurred. The plastic strain is then of the same order as the elastic strain. This yield stress has probably no physical significance but is an arbitrary point taken on a continuous stress-plastic strain curve. Even if a discontinuity can be shown to occur, as in the sharp yield point of mild steel, which is due to the onset of some particular dislocation mechanism, it does not necessarily preclude the existence of other dislocation movements of significance at lower stresses and strain. Figure 2 shows a much enlarged view of the first part of the same curve, using a strain gauge with sensitivity increased from 10^{-4} to 10^{-7}. At very low stress (a), the curve is linear and appears exactly reversible, that is, behaviour is Hookean elastic. At slightly higher stress (b), the loading curve becomes slightly curved over at the top, as if plastic flow was commencing, and on unloading the curve drops initially at the full rate of the elastic modulus; but as the stress nears zero the curve bends in to close the loop. The material can be taken round the loop many times without detectable change, though the work represented by the area of the loop is converted into heat. This effect is called mechanical hysteresis.

Interest in mechanical hysteresis was revived by Roberts, Brown and Ekvall[1,2] and their technique has been widely employed since. The loop area W (work done per unit volume per cycle) is measured as a function of the maximum strain width of the loop ϵ_L. On theoretical grounds it was claimed that they are related by the equation

$$W = 2\sigma_f \epsilon_L ,$$

where σ_f is a friction stress opposing dislocation movement. This is based on the concept of the hysteresis being caused by bowing of dislocations between anchor points. Dislocations have energy and can be considered as being under tension. The applied stress tries to drive the dislocation forward and at equilibrium this is balanced by the tension and a friction stress. On increasing and decreasing the applied stress the plastic strain is different by the strain equivalent of twice the friction stress. Experimental curves of W against ϵ_L have not always been straight lines through the origin and controversy has raged over the correct model. It has recently been shown by Bacon[3] that, if the friction stress is assumed to be orientation-dependent, the equations of dislocation bowing are modified to give a slightly curved relationship between W and ϵ_L which agrees with the experimental data.

Another way of detecting hysteresis is to measure the damping in free or forced vibration. Two new factors are introduced: the stress may not vary between zero and some value; the straining rate or time element. Depending on the material, these factors can be significant. Damping will not be considered further here; the subject has been reviewed by Entwistle[17].

Raising the maximum stress at first increases the size of the hysteresis loop (Fig.2(c)) and then causes the loop to open (d). This is the onset of plastic flow, that is, residual strain on removing the load. Since the magnitude of the strain is so much smaller ($\sim 10^{-6}$) than conventionally detected (10^{-3}), it is called microyielding. Cycling the load to the same stress level now causes further increments of residual strain. As with hysteresis and macroyielding, microyielding is normally detected at a stress level which is determined by the sensitivity of the strain gauge. However, in some materials it has been shown that the stress-plastic strain curve can be extrapolated to zero strain to give a finite intercept stress with physical significance. This will be discussed later.

A final effect is due to the time factor (see Figure 3). Even in the range of elastic hysteresis, the loop has been observed to close at the end of a load cycle only after a period at zero stress. Similarly, at the maximum stress the strain increases slightly with time. Because the strain is linear with log (time) the creep rate falls rapidly. At higher stresses, creep increases both on loading and unloading. Nevertheless total creep strains are small in comparison with the elastic strain and limited with time. Creep is a thermally activated process and thus very sensitive to temperature. It is a mistake in general to try to correlate creep phenomena with microyielding, which is due to stress applied for a short period.

Although all these related aspects of microplasticity are known, there are very limited comprehensive experimental data which seek to co-ordinate the different effects. In the following most attention will be paid to microyielding.

3. MICROYIELD STRESS (MYS)

It has already been mentioned that the conventional yield stress is taken at the point at which the plastic deformation is of the same order as the elastic deformation ($\sim 10^{-3}$). Actually the deviation from the initial straight portion of the stress-strain curve is used, but this is effectively the plastic deformation (defined as residual on removing the load). Over the last thirty years[4] there has been increasing interest in the small-scale deviations from the elastic line, that is, where the plastic strain is $\sim 10^{-3}$ of the elastic strain.

A microyield stress (MYS) has been defined as the stress at which a plastic strain of 10^{-6} develops (the corresponding elastic strain usually falling in the range 10^{-4} to 10^{-3}). Some workers, using less sensitive strain gauges, have used a value of 2×10^{-6} plastic strain, and this will be denoted here as MYS_2. Although the MYS can be established in the same way as the conventional yield stress (YS) by detecting deviation from the initial straight line, a more satisfactory technique in the circumstance where a very small strain is to be detected against a large one is to remove the large elastic strain component by unloading. The load, increased by steps, is applied and removed, and the MYS is then the stress at which a residual plastic strain of 10^{-6} occurs. Slightly different values may be given by the two methods, due to elastic hysteresis effect.

It was claimed initially that the MYS was at last the limit of ideal elasticity or, at the very least, a definite discontinuity in the stress-strain curve. Hughel[5] used the term "precision elastic limit" (PEL) in the belief that the point had a physical significance. He claimed that a step yielding of 10^{-6} in/in occurred at this point in beryllium and that no further yielding occurred until a much higher stress was reached. However, the evidence was not very strong as the sensitivity of his strain gauge was only 10^{-6}. Again, Muir, Averbach and Cohen[6] found with hardened steel that repeated applications of the load to a fixed stress level produced no residual plastic strain per cycle until the MYS_2 was exceeded, when it increased rapidly with stress level. Again the strain sensitivity was about 10^{-6} and it is not clear that there was a definite discontinuity.

The only satisfactory method of showing that an effect really does begin sharply rather than develop progressively is to plot the results with the axes chosen so that the points lie on a straight line. Then an intercept with the stress axis or the intercept of two straight lines of different slope are real proof of a discontinuity. This is not to say that in Figure 1 there is no yield stress below which elastic deformation is dominant and above which plastic deformation is dominant. The question at issue is whether there is a component of plastic flow during the predominantly elastic range. In the work of Bonfield and Li[7] on beryllium it was shown that the microyield curve (stress σ versus residual strain ϵ) obeyed the law

$$\sigma = \sigma_0 + B\epsilon^{1/2}$$

where σ_0 and B are constants of the material. The value of the intercept stress σ_0 was found by extrapolating a straight line back to zero strain. This showed that this particular mode of plastic flow required a finite stress σ_0 to initiate it, but it does not preclude a further mechanism which may be significant at stress and strains below those detectable with the particular equipment. Another example is found in the paper by Roberts and Brown[1] in which plotting log (stress) against log (residual strain) for zinc single crystals gave two strain lines intercepting at a strain of 10^{-4}. This point must clearly be a yield discontinuity not involving the instrument sensitivity or experimenter's judgement.

The MYS has been investigated both for the light shed on fundamental theory of dislocations and for its possible significance in the development of precision instruments. For the latter, there has been a tendency to assume that the MYS is the limit of ideal elasticity and that at stresses below it there will be no elastic hysteresis, no microcreep and no material instability. These are sweeping assertions and it is to be hoped that the preceding discussion will cause the MYS to be seen through less rose-coloured spectacles. It has been shown, however, that the MYS is a fraction of the conventional yield stress and that the ratio of the two varies very widely from one material to another, and with the details of the manufacturing process, including any heat treatment. Results on specific materials will now be reviewed from a phenomenological viewpoint.

4. REVIEW OF RESULTS ON SPECIFIC MATERIALS

4.1 70/30 brass

In early measurements of microplasticity Smith found that 70/30 brass had a conventional yield stress (hereafter called CYS of 28,000 psi (lb/in^2) but the MYS_2 was 23,000 psi. Pre-stretching up to one per cent caused little change in the CYS but dropped the MYS rapidly to 10,000. Further pre-stretching up to 10% raised the CYS to 50,000 psi, leaving the MYS steady (see Figure 4). Annealing at 300°C raised the MYS_2 to 30,000 psi.

It appears that in this material there is a very marked strain softening of the MYS in contrast with the normal strain hardening process. This result was not confirmed in beryllium copper and beryllium (discussed later).

4.2 2% Beryllium Copper

Precipitation hardened beryllium copper is used for its good elastic properties and electric current carrying ability in instrument suspensions, gyroscope ligaments (pigtails) and similar applications.

The MYS_2 was measured by Smith and Wagner[8] in various conditions, following solution treatment (quenching from 825°C). It reached a maximum value of 100,000 psi after cold drawing in the solution treated state and precipitation hardening (3 hr at 300°C). The corresponding CYS (0.1%) was 170,000 psi. In the solution state the MYS was only 10,000 psi, increased by cold working to 44,000 psi. Cold working thus had a contrary effect to that found in 70/30 brass, but in line with normal strain hardening. This may have been due to residual stresses always found in quenched material.

Very much lower values of MYS_2 were reported by Bonfield[9], although the test piece was thin strip (0.18 × 0.06 inch) instead of round bar (0.5 inch diameter). The MYS_2 was only 4500 psi after cold work and precipitation hardening. In the solution treated state, the value was 1600 psi, increased to 3000 psi by cold reduction (40%). The stress-residual strain curves showed three stages: a low initial strain hardening rate, a high one and then nearly zero. Results, reproduced in Figure 5, suggest that precipitation hardening is only effective after 20×10^{-6} strain. Retesting after straining gave only slightly higher values of MYS_2. Dislocation structures and hysteresis curves were also obtained. The large discrepancy with the earlier work of Smith and Wagner[8] was not considered.

4.3 Steel

Steel in the hardened state or hard drawn state is used for springs and bearings, in the normalised state for structural components. The phase changes involve very large dimensional changes: 8×10^{-3} linear contraction on austenizing; 14×10^{-3} linear expansion on quenching to martensite and 6×10^{-3} contraction on full tempering. (Values are for 1% carbon steel.) The changes on quenching and tempering do not usually proceed to completion, but will transform further with time and stress. Large residual stresses are also present. The complex nature of the problem of rationalizing data on microyielding involving strains of 10^{-6} will be apparent.

The effect on the MYS_2 at 20°C of varying the tempering temperature for the steels with carbon contents from 0.2 to 0.8% was investigated by Muir, Averbach, and Cohen[6]. Negative strain up to 20×10^{-6} was found in some specimens and the MYS_2 was taken at the stress above which the residual strain increased continuously in the positive sense. Load cycling above the MYS_2 produced increments of residual strain. Results at all carbon contents showed that the MYS_2 is low for as-quenched martensite, but increased on tempering up to 400°C and then fell. This is quite different from the behaviour of the conventional strength parameters such as CYS, tensile strength and hardness, which are initially high and decrease on tempering. Typical results for 0.82% carbon steel are reproduced in Figure 6. The maximum MYS_2 was almost independent of carbon content, at around 100,000 psi, and the use of high carbon steel in springs must be attributed to the higher CYS which can be obtained. Austempering (isothermal transformation at 305°C) gave lower values of MYS.

Data on the MYS of cold drawn steel wire has not been found. Normal practice is to use about 0.8% carbon and convert to fine pearlite by isothermal transformation of austenite (450°C), a process called patenting. After cold drawing up to 80% reduction of area, a low temperature anneal at 250°C is given to raise the 0.1% proof stress, which tends to be low in relation to the tensile strength, which is up to 400,000 psi.

The microcreep of 1.4% chromium steel (AISI-SAE 52100) as used for rolling element bearings has been investigated by Mikus and his co-workers[10]. Four heat treatments were tried: (i) quench and temper at 127°C (standard practice); (ii) quench and temper at 260°C; (iii) quench and temperature cycle ten times, -196°C to +127°C; (iv) austemper at 260°C (isothermal transformation to bainite). Stability and creep tests (maximum stress 20,000 psi) were done at -34°C, 35°C and 74°C over 1500 hours. Stability varied between $\pm 10 \times 10^{-6}$ over the period, with heat treatments (ii) and (iv) giving near zero. Microcreep was only important at 74°C with the strains approaching 100×10^{-6} for (i) and (iii), and 10×10^{-6} with heat treatments (ii) and (iv). The presence of retained austenite was considered the main source of instabilities and austempering was recommended. This was disputed in discussion as producing too low a hardness, leading to high wear rates in bearings. Normal hardness is 64 R_c (800 DPN) and austempering gave only 57 R_c (635 DPN).

The microyielding of normalised mild steels has been investigated by Hahn and others[11] in connection with the problem of brittle fracture. The MYS_2 was coincident with the lower yield stress in fine grained material, around 40,000 psi, but slightly lower than it (18,000 psi of 22,000 psi) in the same material heat treated to a coarser grain size.

4.4 Beryllium

Beryllium is the principle structural material in precision gyroscopes because of its low density, high elastic modulus, and a temperature coefficient of thermal expansion matched to steel.

Its microplasticity has been extensively explored on this account and in the pursuit of basic research by Hughel[5,12] and by Bonfield and Li[7,9,13-15]. A short paper by Ruckman and White[14] describes some work at the United Kingdom Atomic Energy Authority, Aldermaston. Hughel measured MYS at 62.8°C (and some microcreep) whilst Bonfield measured MYS_2 at room temperature. Using the stress-residual strain law found by Bonfield[4], the difference between MYS and MYS_2 is about 400 psi at room temperature. Variation with temperature has not been explored explicitly, but it is probably small and a net difference of 500 psi is about right. This is small in hot pressed electrolytically-polished beryllium with MYS of 11,000 psi, but significant with annealed specimens with an MYS in the range 750-3000 psi.

The main conclusions of the work of Hughel and Bonfield are as follows:

(1) Standard QMV grade (from the Brush Beryllium Company) has a yield stress (0.2%) of 35,000 psi and a MYS of 3500 psi when annealed after final grinding (2 hr 800°C)[5].

The MYS of hot pressed electrochemically polished QMV is 11,000 psi (Refs.7 and 13).

(2) The scatter of results is large. The block to block variation in the annealed condition is 1500 psi, and within a block 800 psi (Ref.5).

(3) Surface damage due to machining lowers the MYS_2 of hot pressed QMV from 11,000 psi to 6000 psi; electropolishing away 0.01 inch of the surface restores the value[13].

(4) Annealing (2 hr 800°C) also removes surface damage but alters the dislocation structure from a high density with tangled arrays to low density with dispersed arrays. The MYS_2 falls from 10,000 to 3000 psi (Ref.7).

(5) The drop of MYS due to surface damage is caused by twins acting as stress raisers rather than the residual stresses[13].

(6) The microyield curve obeys the law

$$\sigma = \sigma_s + B\epsilon^{1/2} ,$$

where σ_s and B are material constants. The change of MYS due to annealing is due to change of σ_s (from 8000 to 800 psi) no. of the strain hardening rate B (Ref.7).

A transition in the yield curve occurs at higher strains ($\simeq 10 \times 10^{-6}$) (Ref.14).

(7) Hysteresis loops were detected above 1000 psi in annealed metal. The friction stress (which is independent of the strain sensitivity) was $\simeq$ 400 psi and increased to $\simeq$ 4000 psi with small pre-strains (100×10^{-6}) and then remained almost constant. For hot pressed beryllium the friction stress was $\simeq$ 5000, independent of pre-strain[15].

(8) Microcreep tests (called "dimensional stability" (DS) by Hughel) carried out at 95% MYS, 62.8°C, showed extensions of $\simeq 6 \times 10^{-6}$ in 500 hr (Ref.5).

(9) High oxide and fine grain size raises the MYS from 3500 to 13000 psi for metal annealed after final grinding (Ref.5).

5. NEW MEASUREMENTS OF MYS OF BERYLLIUM

An investigation of the microplasticity of beryllium from the Royal Ordnance Factory (ROF) at Cardiff is now taking place at the British Aircraft Corporation, Stevenage, in conjunction with the Royal Aircraft Establishment, Farnborough. The work arose out of some alarming examples of dimensional instability of gyro components some five years ago which, it is now realised, were caused by unsymmetrical machining stresses rather than inherent disabilities of the beryllium.

Since the objective is to build precision components, rather than pursue the delights of metal physics, the test specimens have usually been machined similar to those possible in practice: stress relieved for 1 hour at 800°C before final turning of 0.0002 inch surface. Fuller details will be published shortly. The results so far obtained, which are almost identical with that found on Brush material, are as follows:

(1) Standard grade ROF beryllium (2.3% BeO) has a MYS of 1070 psi when annealed and electropolished (6 specimens). Electropolishing did not have a significant effect, but only 0.0003 inch was removed.

(This is a factor of two down on Bonfield's results[7] on QMV materials, allowing 500 psi difference between MYS and MYS_2. He removed 0.010 inch before the MYS_2 on hot pressed material increased from 6000 to 11,000 psi.)

(2) The MYS is 1040 psi when stress relieved before final turning (36 specimens). Tests on Brush QMV grade obtained for comparison gave 1270 psi (22 specimens).

(3) With no heat treatment the MYS is 1310 psi (4 specimens).

No polished unheat treated specimens were tested, which Bonfield found to have a very high MYS of 11,000 psi. It would be interesting to know why annealing has such a large effect on polished specimens as hot pressing appears to involve annealing during the cooling period.

(4) The standard deviation from block to block (7 blocks, 4 specimens each) is 100 psi. Due to skew distribution about the mean this figure does not fully reflect the actual scatter. Within a block the standard deviation was 40 psi.

(5) The stress-residual strain curve obeys Bonfield's relation

$$\sigma = \sigma_c + B\epsilon^{1/2} .$$

Typical results are plotted in Figure 7. The values of σ_c fall between 0 and - 500 psi.

(The negative values are possibly due to the surface damage present. No hot pressed material with polished surfaces was tested, which Bonfield found to have very high σ_c, around 8000 psi (Ref.7).)

(6) High oxide material (5.8% BeO) has a higher MYS of 2110 psi (cf. 1040 psi) (4 specimens).

Tests on Brush 1400 grade gave 4150 (3 specimens). This difference is attributed to ROF material in this test being pressed from 200 mesh powder whereas the Brush material is 350 mesh or better.

6. LIGAMENT CREEP

Ligaments are used both in gyroscopes and accelerometers to carry currents to the sensitive element or pendulous arm. The mechanical restraint has to be kept as small and as stable as possible. Recent experience with a precision gyroscope has shown that microcreep in ligaments can be a source of variable drift just after switching on and can be a cause of delay in aligning the inertial navigation system.

The effect was observed in a single-degree-of-freedom gyro under test in the torquer feedback mode with the output axis vertical and input axis pointing East. Figure 8 shows the torquer current (converted to equivalent drift rate) to hold at null after the sensitive element had been allowed to rest on one or other of the stops (set at a nominal $\pm 2^{\circ}$) for 2 hr or 16 hr. The gyro was kept running, so that no variability due to warm-up appears, but the recorder was connected in at the start of the test. This introduces a small transient which can be seen in curve 4 for which the gyro had been on null and was then connected to the recorder.

Figure 9 shows similar curves for the same type of gyro but with the stop freedom reduced to $\pm 0.5^{\circ}$ and the ligament material changed from 85-15 silver copper alloy to 62.5% gold-copper-silver alloy.

There are seven hairpin-shaped ligaments formed from strips 2.5×10^{-4} inches thick by 4 or 7×10^{-3} inches wide. They introduce a torque on the float of 0.62 dyn cm ($= 0.36^{\circ}$/hr drift rate) per degree of float rotation for the silver alloy and 1.5 times this value for the gold alloy on account of the higher elastic modulus. The stress varies along the ligament, with a maximum value of 460 psi at the anchorage clips, to which they are soldered. The ligament strip is received cold worked and is formed into ligaments, annealed (1 min at 500°C) and then quenched. The MYS and hysteresis behaviour is not known, but work is progressing to further study and reduce this effect.

ACKNOWLEDGEMENT

This paper is Crown Copyright and is reproduced by permission of the Controller, Her Majesty's Stationery Office.

REFERENCES

1. Roberts, J.M. Brown, N.	Transactions, Metallurgical Society, American Institute of Mining Engineers. Vol.218, 1960, p.454.
2. Brown, N. Ekvall, R.A.	Acta Metallurgica. Vol.10, 1962, p.1101.
3. Bacon, D.J.	Acta Metallurgica. To be published.
4. Smith, C.S.	American Society for Testing and Materials. Vol.40, 1940, p.864.
5. Hughel, T.J.	Institute of Metals Conference on Metallurgy of Beryllium, London, 1961.
6. Muir, H. et al.	American Society for Metals. Vol.47, 1955, p.380.

7. Bonfield, W. Li, C.H.	Acta Metallurgica. Vol.11, 1963, p.585.
8. Smith, C.S. van Wagner, R.W.	American Society for Testing and Materials. Vol.41, 1941, p.825.
9. Bonfield, W.	Transactions, Metallurgical Society, American Institute of Mining Engineers. Vol.239, 1967, p.99.
10. Mikus, E.B. et al.	American Society for Metals. Vol.52, 1960, p.307.
11. Hahn, G.T. et al.	Institute Conference on the Atomic Mechanisms of Fracture, Swampscott, 1959. Massachusetts Institute of Technology, John Wiley.
12. General Motor Corporation	*Further Results of an Investigation of the Precision Mechanical Properties of Improved Forms of Beryllium*, 1962.
13. Bonfield, W. et al.	Transactions, Metallurgical Society, American Institute of Mining Engineers. Vol.227, 1963, p.669.
14. Bonfield, W. Li, C.H.	Acta Metallurgica. Vol.12, 1964, p.577.
15. Bonfield, W. Li, C.H.	Acta Metallurgica. Vol.13, 1965, p.317.
16. Ruckman, C. White, J.	Conférence Internationale sur la Métallurgie du Beryllium. Grenoble, 1965, Presses Université de France.
17. Entwistle, K.M.	Metallurgical Reviews, Vol.7, 1962, p.175.

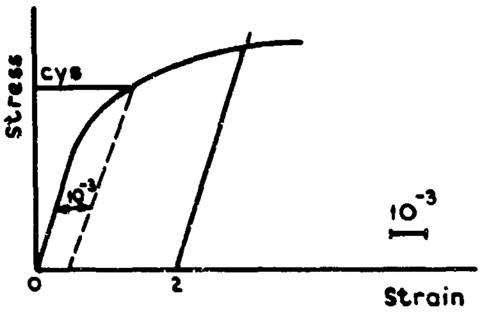


Fig.1 Macroyield

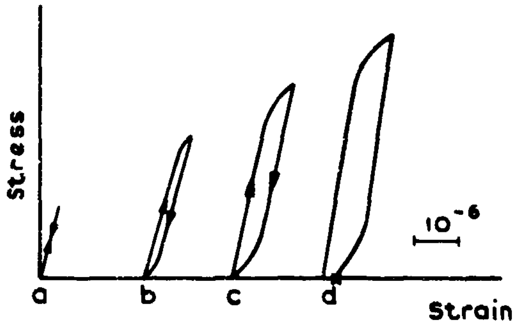


Fig.2 Elastic hysteresis and microyield

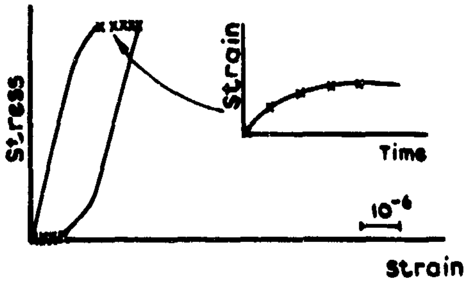


Fig.3 Microcreep

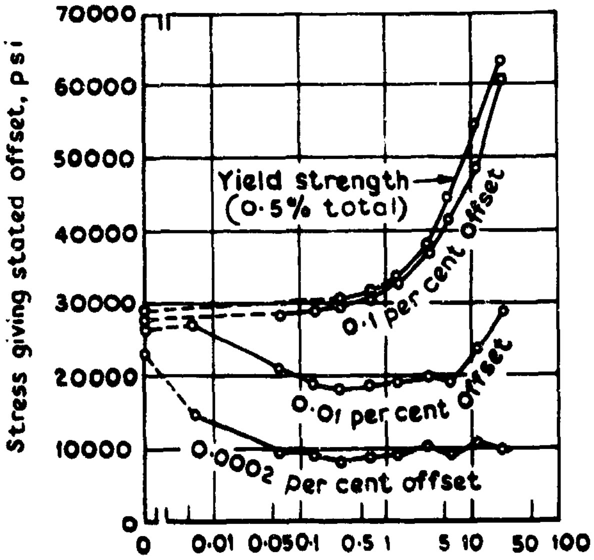


Fig.4 Copper. Effect of prior stretching on stress corresponding to various amounts of offset from proportionality (Smith[4])

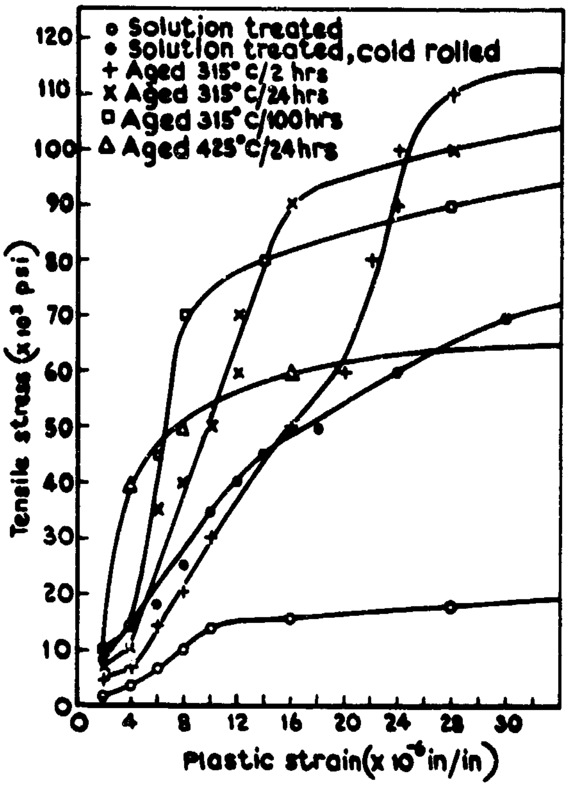


Fig.5 Beryllium copper (Bonfield[9]). Stress-strain curves in the microstrain region as a function of cold work and ageing

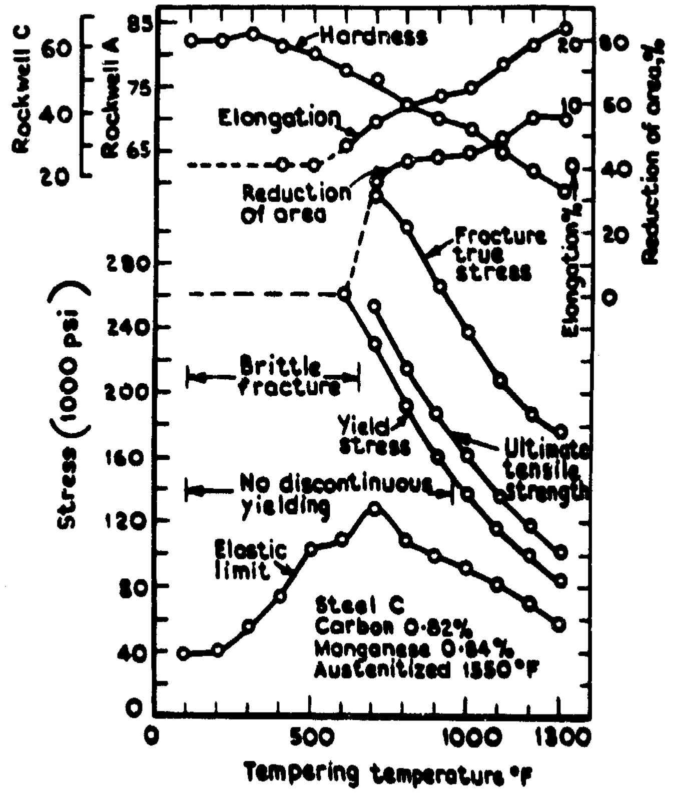


Fig.6 Hardened steel (Muir, Averbach, Cohen[6]). Average mechanical test results for 0.82% carbon steel

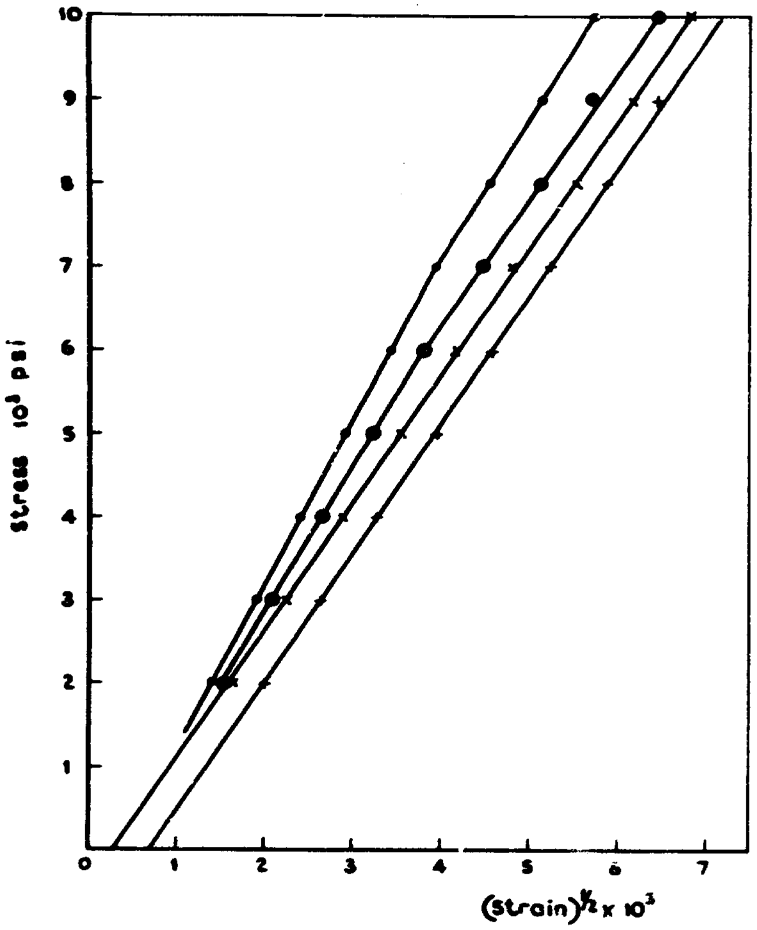


Fig.7 ROF beryllium. Stress versus (residual strain)$^{1/2}$ with normal heat treatment

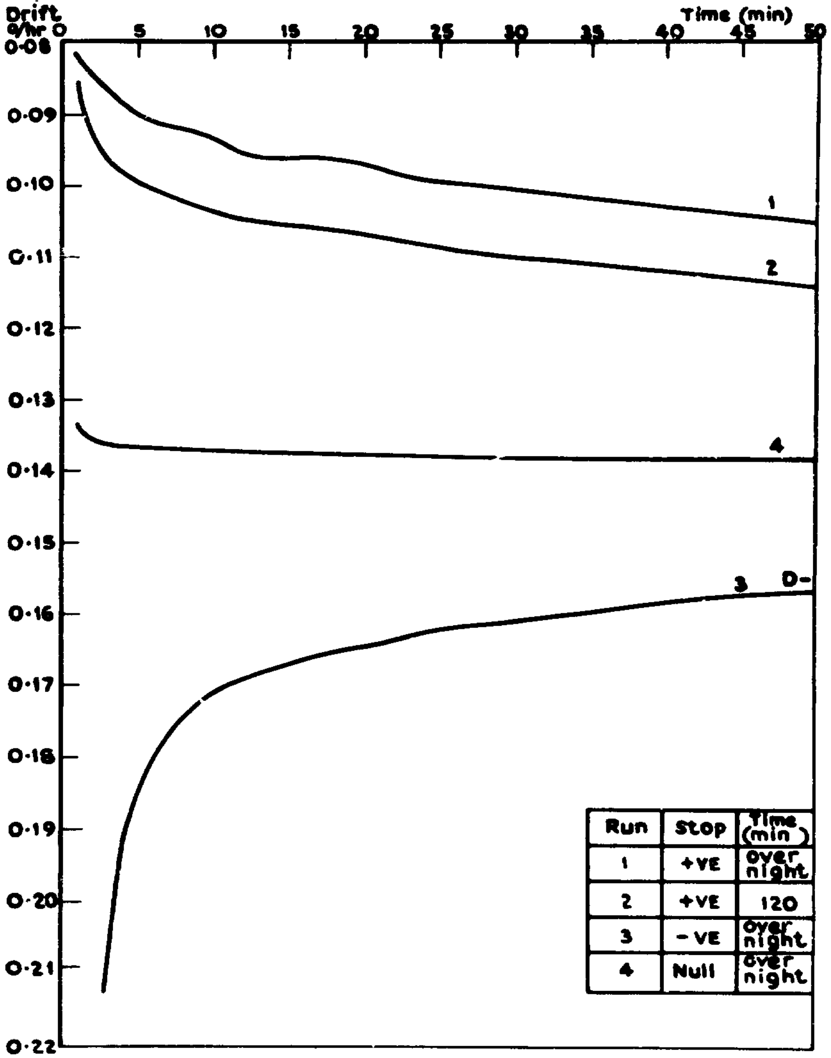
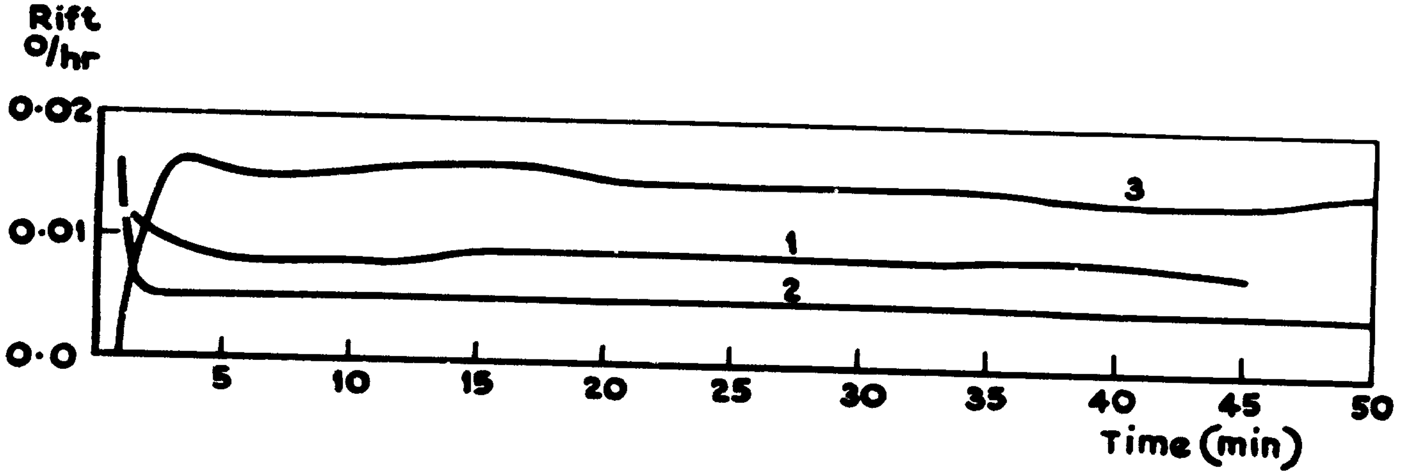


Fig.8 Ligament creep effect: gyro drift following a period held on the stops



Run	Stop	Time (min)
1	+ve	over night
2	+ve	120
3	-ve	over night

Fig.9 Ligament creep with improved ligaments

GYRO BALL BEARINGS - TECHNOLOGY TODAY

by

Albert P. Freeman

Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

SUMMARY

The spin-axis bearing package is a major determinant of gyro reliability and performance. Fractional microinch position stability over extended time periods, reliably achieved for tens of thousands of running hours, is required. Whether the ball bearing thus used runs successfully on an elastohydrodynamic fluid film or succumbs to early failure may be determined by whether or not today's bearing technology is applied. This specialized technology, applicable in many aspects to other bearings, has been developed over the past twenty years and continues to advance.

Achievement of current state-of-the-art is the result of parallel development of the bearing parameters and of the means for their evaluation. Bearing metallurgy, geometry, groove-surface topography and chemistry, lubrication, ball-retainer, contamination control, dynamic behavior, testing, and processing variables have all been improved. Particularly significant have been the efforts in surface-film-piercing asperity reduction, surface-chemistry improvement, and lubrication-mechanism advancement. Also of major importance has been development or adaptation of measuring devices to join with functional tests in evaluation of bearing characteristics and of potential life and performance at various processing stages.

Continued current effort in the areas of the lubrication mechanism, bearing dynamics, and groove surface promise further gains in consistency of achievement of life and performance goals. Application of today's technology can in most cases, however, yield the required thousands of hours of reliable operation.

GYRO BALL BEARINGS - TECHNOLOGY TODAY

Albert P.Freeman

1. INTRODUCTION

The gyro spin-axis ball bearing is unique. At the heart of the inertial guidance system, it is a major determinant of performance and reliability. Today's bearing technology is the product of more than twenty years of development, though current practice is in many cases frozen at a point dating back many years. Use of the knowledge available now can improve bearing yield, performance, life, and reliability.

1.1 Requirements

The principal requirement of the inertial gyro spin-axis ball bearing can be stated very simply: long life at the required performance level. We can divide this requirement into two broad subcategories: freedom from physical or chemical degradation of all elements of the bearing package and maintenance of dimensional stability of the gyroscope element. Depending upon the application, failure criteria can range from slight deterioration of gyro performance to inability of the wheel to turn because of bearing seizure.

This paper is concerned with bearing performance in precision floated inertial gyros of the type shown in Figure 1, and with those factors that influence stability of the gyro wheel package. As indicated by Figure 2, average wheel location must be stable to a fraction of a microinch, and mass stability of other float elements must be similarly very closely maintained.

In order to satisfy the stability requirements, the bearings must be supported on a full, stable elastohydrodynamic (ehd) lubricant film. Piercing of this film during running causes chemical and physical degradation in the lubricant and of the metal surface, which in turn influences the location of the gyro wheel. This sort of deterioration is progressive, as the debris formed by the high-speed metallic contact leads to further piercing of the film. The lubricant sludge then collects beside the pressure zones and, like a sponge, withdraws the oil from the region in which it is needed to maintain the film. As this mode of failure progresses, bearing torque becomes erratic, the lubricant vanishes, the metal wears, and ultimately the torque increases to the point at which the wheel will no longer run at operating speed. The bearing running time between the onset of performance degradation and wheel failure can be several thousand hours. It is interesting to note that metal fatigue, one of the classic modes of bearing failure, plays essentially no part in gyro bearing failure.

Generation and maintenance of the ehd film demands the continued existence of many conditions. In operation, the metal components (races and balls) must have a geometric form that generates the required ehd film with acceptable stress levels over the entire pressure zone. The metal must sustain the load essentially without plastic flow or surface damage. The surfaces must be free of film-piercing asperities and must chemically support a boundary lubricating film at low speed and an elastohydrodynamic film at operating speed. The lubricant must demonstrate the chemical and physical properties needed to achieve these films with acceptable torque levels, along with chemical and thermal stability. The ball retainer must maintain a controlled lubricant reservoir and circulate this lubricant as needed for a full ehd film, and run with required stability at acceptable torque levels. The environment must be chemically, physically, and thermally compatible with the bearing package. Finally, and of extreme importance, the bearing package must be free of contamination that can cause bearing degradation due to piercing of the film.

In addition to the demands just noted, which are associated primarily with retention of the lubricant film, the bearing package must also demonstrate other properties needed for mass stability. The combination of geometry, lubrication, and operating parameters must assure such factors as

(a) relative insensitivity of wheel location to acceleration field variation,

(b) film uniformity and stability of the bulk lubricant,

(c) stability of ball group and retainers,

(d) constancy of bearing torque.

Inadequacies in the first group of parameters discussed, which are associated with rupture of the ehd film, result in bearing deterioration and gyro performance degradation. The second group, associated with mass instability, generally influence instrument quality without necessarily reducing bearing life at the degraded performance level.

Of utmost importance in any discussion of bearing requirements is processing, or handling. Achievement of the basic bearing properties is vitally dependent upon quality control during manufacture, which can be accomplished with proper engineering supervision. Retention of these properties is then a major battle. The bearing is a precise device and must be treated as such. During processing from package to completed instrument, its integrity can be compromised by particulate contamination, chemical contamination, overheating, mounting distortion, overstressing, scratching, denting, shock, overlubrication, underlubrication, exposure to corrosive environments, etc., etc. Thus, the processing variables are extremely important in both initial achievement of required bearing parameters and in their retention during instrument fabrication.

1.2 Status

The requirements noted in Section 1.1 have been achieved. They can be achieved consistently. Gyro performance of the highest quality has been demonstrated in instruments that have accumulated about 30,000 wheel running hours. However, as will be noted in subsequent sections, further advances can still be made in some areas of bearing surface, lubrication, and dynamics.

Achievement of today's status results from more than twenty years of development work, some phases of which are still in progress. Figure 3 illustrates this work. One of the first major development efforts, initiated in the late 1940's, led to improved preloading techniques. Subsequent efforts in the bearing package have encompassed metallurgy, geometry, lubrication, retainers, surface finish, surface chemistry, dynamics, contamination, manufacturing techniques, and processing variables.

Progress in the field of bearing evaluation has been following a parallel path of equal importance. For example, the early preloading improvement noted above was accompanied by development of an axial yield gauge. Other developments include improved geometric measurement devices, optical measurement techniques, the low-speed dynamometer, the milliwattmeter, the lubricant-film electrical-resistance gauge, taper sectioning, stroboscopic observation, race counter-rotation devices, high-speed torque testers, and many others.

The gyro itself is one of the most useful bearing diagnostic devices. It alone is capable of determining bearing position stability to the required performance levels. It also provides a convenient means for the application of known inputs to the bearing package, along with precise readouts of the accompanying bearing behavior.

Bearing package evaluation in early gyro construction and test stages is extremely important. The ball bearing is unfortunately quite forgiving on a short-term basis, and early degradation symptoms are frequently ignored. Progressive deterioration can lead to later severe performance degradation, perhaps when the gyro is the heart of a complex, critical, operating navigation system.

The following sections of this paper will discuss the current status of gyro ball-bearing technology, with particular emphasis on areas known to be critical. No attempt will be made to cover the history of the developments that have led to today's status, except where required for perspective. Due to the complexity of the subject, the treatment will be limited in this paper and discussion of each subject will be relatively brief. References providing further details are included in the bibliography.

2. CONFIGURATION

Gyro bearing design must consider not only the normally accepted criteria but also those peculiar to the precision gyro, such as microstability, isoelasticity, lubrication limitations, bearing dynamics, and long-term physical and chemical stability. These factors, coupled with specific instrument requirements and configurations, yield the basic bearing design. Among the design features are basic size and configuration, materials, mounting method, preload, speed, lubrication, contact angle, race-groove curvature (groove-to-ball conformity), inner- or outer-land relief, retainer configuration, and many others. (The bearing nomenclature used here is defined in the Appendix on p. 330). Tolerances must also be assigned to most of these parameters.

Metal and geometry are discussed in this section, with particular emphasis on adherence to nominal values, or tolerance control. Design criteria leading to the specific configuration are not discussed because of their complexity and dependence upon the details of the requirements of the gyro in question. For example, depending upon the acceleration environment to which the gyro will be exposed and the performance demanded during acceleration (vibration or steady-state), critical bearing parameters may be contact angle, number of balls, race-groove curvature, and preload. On the other hand, torque limitations may emphasize basic size, speed, preload, race-groove curvature, atmosphere, lubrication, and retainer. As these brief examples point out, bearing design is critical, but it is too complex a subject for coverage here. The influence of specific geometric variables on lubrication and on bearing dynamics will be discussed in later sections.

2.1 Metal

The demands made upon the steel of the gyro ball bearing differ somewhat from those made on other more heavily loaded bearings and are in some respects more severe. The gyro bearing typically is lightly loaded, with the maximum Hertz stress generally in the vicinity of 300,000 lb/in^2 or less. It operates in a moderate ambient temperature, about 150°F in an inert atmosphere, generally helium, after having been very carefully processed from production through application. Why, then, are we concerned with the properties of the steel from which the bearing is made?

The steel must satisfy two major requirements. First, the level of gyro performance as discussed in Section 1.1 demands the ultimate in microdimensional stability, both under stress and unloaded, as well as over a wide temperature range such as a -65°F to +225°F range of in-process thermal cycling and storage. Second, the microstructure must be such as to permit the ready generation and maintenance of race-groove and ball surfaces physically and chemically capable of both boundary and elastohydrodynamic (ehd) lubrication under the unique running conditions of the gyro bearing. In this regard, freedom from ehd-film-piercing asperities is extremely important.

Both 52100 and 440C, the most commonly used gyro bearing steels, have demonstrated the ability to meet these requirements. Other steels have also been used successfully in gyro bearings, such as M-2, M-50, and WB-49. The steel most commonly used in gyro bearings through the years has been, by far, 52100. It has been quite satisfactory, but recurring unpredictable instances of corrosion have presented a problem. In the past several years, 440C has been attaining greater popularity for its resistance to corrosion and because of successful application. Both steels are readily fabricated to the required geometry and surface finish, in spite of the difference in microstructure primarily caused by the relatively large carbides in 440C, as seen in Figure 4. In controlling the steel, factors of concern include the following: chemical composition; microstructure; carbide type, size, and distribution; response to heat treatment. The last, in turn, encompasses microstructure, strength, hardness, retained austenite, corrosion resistance, and stability.

Specific precautions are still warranted in the selection and application of both 52100 and 440C, in spite of the demonstrated ability of both of these steels to yield successful gyro bearings. These precautions include the assurance of freedom of the steel from nonmetallic inclusion and control of processing variables concerned with heat-treatment and metal-removal operations in the hardened state.

The presence of inclusions in other types of bearings which are highly stressed is detrimental because inclusions provide initiation points for fatigue failure. In gyro bearings, inclusions are also a serious problem, but for a different reason: they limit the surface achievable for the generation of a full hydrodynamic film, are associated with film-piercing asperities, and can cause chemical and particulate contamination problems. These effects will be covered more fully in Section 3; this section is more concerned with recognition of the problem in the steel.

Bearings with poor surfaces and low manufacturing yield, both attributable to inclusions in the steel, occur in spite of inspection of the steel for cleanliness by accepted rating methods, e.g., the JK (Jernkontoret) method. Stringers in the steel are particularly difficult to detect by conventional means. One approach to an improved steel-rating method is to examine steel surfaces in regions of the bar more representative of the bearing than the small flat sections normally examined. This is done by machining sample bars with steps at successively smaller diameters, representing diameters of the race-groove functional surfaces. These steps are then honed and examined microscopically for inclusions, as seen in Figure 5. Steel lots evaluated by both this technique and the conventional approach have shown the honed step-down bar evaluation to correlate far better with race-groove surface topography and manufacturing yield.

Heat treatment of the bearing parts plays a major role in establishment of the previously noted properties of the finished bearing. It is important not only to determine optimum heat-treatment parameters but also to assure rigid adherence to the values selected. For this, the testing of sample pieces is needed.

During austenitization, critical control often is required of the atmosphere and of the temperature level and timing cycle. Important to quenching are temperature, timing, oil-bath cleanliness, and agitation. Subcooling and tempering demand control of timing, temperature and, in some cases, medium. Tests conducted to assure quality may include, as required, hardness, metallographic examination for surface modification and microstructure as shown in Figure 6, retained austenite, and dimensional stability.

Properly selected, tested, and processed, today's steels are capable of the most rigid performance requirements demanded of current gyros.

2.2 Geometry

Nominal bearing geometry and variations therefrom play a major role in establishing gyro life and performance. With regard to life, geometry influences stress levels, thickness of the ehd film separating balls from races, and lubricant control. With respect to gyro performance, geometry influences bearing dynamics, lubrication stability, and response of the gyro to acceleration fields. In addition, geometric tolerance levels influence gyro producibility.

One of the early major efforts leading to present gyro bearing technology was the development of measurement tools and techniques and then fabrication methods to achieve an order-of-magnitude improvement in geometric tolerances, from typical 0.0002-inch values to levels of 20 microinches. Gyro producibility immediately benefited by this improvement, as bearing-to-wheel and bearing-to-shaft fits became achievable on a tolerance rather than a selection basis. In addition, preloading certainty improved, as did bearing dynamic behavior. Bearing life itself, however, was not significantly affected until race-groove geometry was further improved, along with advances in groove surface characteristics, as is noted in Section 3.

It is quite apparent that the ehd film thickness is influenced by local stress levels, which in turn are affected by race-groove runout and cross curvature. Groove runout itself is the product of various geometric parameters, including roundness, lobing, groove-to-face and face-to-face parallelism, concentricities, face-to-bore and face-to-OD (outside diameter) squareness, and mounting distortions. Bearing manufacturing technology has advanced to the point that runouts can be held to the 5-50 microinch region. Mounting dimensions and forces must be carefully controlled, however, to prevent significantly greater runout due to distortion.

Race-groove cross curvature must also be closely controlled. Thickness of the ehd film in the pressure zone is a function of the local cross curvature. It is important to guard against excessive breaking of the corner at the conjunction of the groove and land. The resultant rounding or chamfer encroaches on the pressure zone and affects the local curvature, in some cases nonuniformly around the race. Figure 7 illustrates the results of this edge rounding.

Development of the Talyrond and a number of other roundness-measuring machines has made reliable measurement of race-groove runout and cross curvature possible. Thus, the generation and measurement of geometry to levels required today is within the capability of current bearing technology.

3. SURFACE

Generation and maintenance of the needed ehd film is a function of race-groove surface topography and chemistry. The surface must be free of film-piercing asperities and must chemically support a lubricating film. Low-speed boundary lubrication similarly depends upon these factors.

3.1 Topography

Shiny or dull, smooth or grooved? This question concerning optimum race-groove surface topography has been one of the most frequently debated for years. The smooth-surface camp wants to maintain the maximum possible ehd-film spacing between opposing surface peaks, by reducing hill-to-valley height, while the striated-surface devotees reason that the valleys between the heights provide a lubricant reservoir. Each group cites convincing data as the basis of its own cause and offers various additional reasons for the superiority of one type of finish to the other.

Actually, both types of surface have operated successfully for many thousands of hours under the most rigorous gyro-bearing running conditions. On the one side have been mirror-like race grooves generated by first running the bearings heavily loaded and submerged in ethylene glycol; surface-finish readings of these bearings were less than 0.3 microinch. Coarse-finish lapped bearings with 3-microinch surface-finish readings have also run very successfully. Smooth surfaces generated by ball lapping and by honing have also fared well. Examples of these finishes are seen in Figure 8.

Within reason, average surface finish does not appear particularly significant. Individual asperities that project above the average surface can, however, pierce the ehd film and bring about failure. Returning to the need for a coarse finish to maintain a lubricant reservoir, any surface that will wet properly with the oil will hold a sufficient thickness of lubricant to permit generation of the required 5-15 microinch ehd film.

Let us examine, then, the factors that should influence the selection of the race-groove finish. Achievement of the required geometry, circumferentially and across the race groove, as well as generation of surfaces that are free of asperities are the two major factors. But we must also be concerned with particulate and chemical contamination, surface integrity (including freedom from "smear"), the "lay" of the finishing marks, ease of inspection and economics. Some bearing race-groove finishing techniques are illustrated in Figure 9.

Perhaps the most commonly used finishing technique is lapping with an abrasive on a string, tape, paper, replica, or other backing. This approach generally improves as-ground roundness, but tends to degrade the cross-groove geometry. It typically produces a striated surface and reacts quite sensitively to irregularities in the metal or particulate contaminants by forming comets, as seen in Figure 10. The grooved surface texture camouflages raised comets and other peripheral asperities, thus making inspection for these features more difficult and expensive. The striated finish, if accompanied by good geometry and freedom from asperities, contamination, and other deleterious factors, has been shown to yield very long successful life. An example is seen in Figure 11.

Honing the groove with a reciprocating, shaped abrasive stone is another common finishing method. If the process parameters are properly controlled, honing yields excellent geometry. Finish depends upon the cycle, choice of abrasive materials, and honing fluid. This process can also generate raised asperities, and the surface finish can range from striated to nearly bland. Figure 12 shows successive improvements in finish accompanying development of improved honing techniques. This process can also yield excellent bearings.

Ball lapping is a newer process that involves lapping the race-groove by an abrasive slurry and groove-conforming balls driven by a rotating cone. This procedure does not improve on the initial race-groove roundness, but it yields excellent cross curvature, characterized in some cases by an omega (ω) shape whose central rise is controlled to keep it out of the pressure zone. Ball lapping yields a uniform matte finish, as seen in Figure 13, and does not generate raised asperities. This natural freedom from certain asperities significantly eases inspection problems. Bearings finished by this technique have demonstrated excellent yield and life.

Other finishing and run-in techniques have been used with varying degrees of success. One experimental approach, prerunning with special fluids, is worth noting for its demonstration of the ability of a bearing with mirror-like race-groove surfaces to run successfully. Running heavily loaded bearings at relatively low speed while submerged in recirculating filtered ethylene glycol will generate very highly polished race-groove surfaces. Such bearings, subsequently tested under gyro operating conditions, have demonstrated long successful life.

Another interesting prerunning technique is that performed in hot TCP (tricresyl phosphate). The resultant surface appearance is only slightly changed, but the bearings demonstrate the beneficial effects of TCP coating discussed under Chemistry in Section 3.2. In addition, bearing yield and life can, under certain conditions, be dramatically improved by this method. A group of bearings made from metal with a high inclusion content and conventionally lapped showed a high incidence of comets. This group also demonstrated low yield and short life. Several pairs of these bearings were TCP prerun, and their yield and life were very dramatically improved. Reduction in the frequency and severity of raised asperities is believed to be the principal reason for this remarkable improvement.

The importance of surface topography, which motivated the work on improved finishing methods, has also led to significant developments in the area of surface-finish evaluation. Electromechanical surface-finish measuring devices have been improved, as have the techniques for application of light and electron microscopy. The simple interference microscope has been particularly useful, as seen in Figures 10, 13, 25 and 26.

One interesting technique for surface topography evaluation is the lubricant-film electrical-resistance gauge shown in Figure 14. This device provides for loading a lubricated ball against a rotating race groove that drives the ball. A ball-to-race electrical circuit provides a measure of asperity contact or conjunction by counting the number of occurrences per revolution of drops in the electrical resistance below a pre-set level. Specific surface features can also be evaluated on a cathode-ray tube. Load, speed, and lubricant are varied. This device is limited by the electrical conductivity of the surfaces and asperities, and chemical coatings on the surfaces. Its use is generating further insight into race-groove surface topography.

Another technique useful in surface and immediate subsurface evaluation is taper sectioning of races, illustrated in Figure 15. A race groove is electroplated for edge preservation, and a chordal sector ground off at a shallow angle. The resultant section is polished and etched to provide a mechanically magnified (by virture of the taper) race-groove surface contour. The metal microstructure close to the surface can also be evaluated by this technique, and microhardness readings can be taken. Such readings typically show that the metal close to the surface is slightly harder than the bulk of the race.

Because of the correlation between bearing life and surface topography, this factor has been improved in many aspects. Today's bearing technology does not have to be limited by surface-topography inadequacy.

3.2 Chemistry

Bearing metal surface chemistry has for many years been cited as a possibly significant determinant of bearing performance. For several years, surface chemistry has been deliberately modified to improve the boundary lubrication capability of the bearing. It has been only recently determined, however, that inadvertently applied chemical surface modifications can adversely affect bearing life under both boundary and ehd conditions.

It has been shown that a lubricated untreated 52100 or 440C bearing will suffer lubricant degradation and surface distress in a running period of less than one hour to several hours at one rpm under normal load conditions. Another interesting phenomenon associated with the boundary lubrication condition experienced at very low speed is the large difference in bearing torque among apparently identical bearing batches received at various times. Bearing torque at one rpm may vary by a factor of three from batch to batch.

Both of these conditions can be corrected by a very simple expedient: prolonged hot soaking of the metal components of the bearing in TCP; the effects of this can be seen in Table I. Life at one rpm then increases from one or a few hours to several hundred or thousand hours. One-rpm friction torque of various batches of bearings then group close together at the low level. It is thought that beneficial effects result from chemical reaction of the acid phosphates present as impurities in TCP with the steel surface. Nitric-acid passivation of 440C steel surfaces has also yielded low-speed life longer than that achieved with untreated surfaces. Pre-running of bearings in TCP, as described in Section 3.1, also produces the beneficial effects described above.

Evidence of detrimental surface chemical modification, or contamination or "poisoning", is more recent and of potentially very great significance. The problem was first recognized when two groups of bearings, which by all conventional evaluation techniques were considered excellent, demonstrated early atypical failure under both boundary and ehd running conditions. These bearings also showed strangely modified surfaces when prerun immersed in TCP, as discussed in Section 3.1. Figures 16, 17, 18 and 19 illustrate these phenomena.

The most effective means for recognition of this contamination was found to be the rate of oil-drop spreading on the race-groove surface; contaminated bearings showed poor oil wettability. A spreading test and typical results are shown in Figure 20. Investigation led to discovery of the probable cause of the poisoning, and its correction led to the delivery of the remaining bearings from one of the two groups in an uncontaminated condition. These "clean" bearings have been used very successfully in a gyro build program, thus further supporting the thesis.

It is important to note that poor wetting of the surface is a symptom pointing to the presence of this contamination, not necessarily an explanation of why early failure occurs. For example, the bearing surfaces can be made to wet with the oil by any of a few techniques, such as immersion in oil or deposition from a solvent solution, and once wet the oil does not spontaneously retract from the surface. An oil drop then applied to a wet surface will spread quite rapidly. Most bearings are used in this prewetted condition. The explanation for failure may lie in the difference in the lubricant properties in the high-pressure zone, particularly the behavior of the molecules next to the surface.

By special solvent-cleaning techniques, a number of the poisoned bearings were rendered "wettable". These bearings are demonstrating greatly improved life under both boundary and hydrodynamic running conditions. An interesting facet of this investigation is the apparent validation of an occasionally reported beneficial effect, derived from aging of bearings stored in oil, and an association of this effect with surface chemistry. Briefly, some of the poisoned bearings have been made wettable by artificial aging (elevated-temperature soaking) in oil, as shown in Figure 21.

It is interesting to speculate on the possible significance of surface chemistry in the age-old problem of unpredictability of bearing-batch behavior: early failure and low yield versus long successful life from batch to batch, with no known difference in the bearing or its application. Current efforts in this investigation are aimed at establishing the fundamentals concerned with the effects noted, improved recognition, prevention, rehabilitation, and means for specifying required surface chemistry. Rudimentary recognition techniques are known today, and means for corrective or preventive action are at hand.

4. LUBRICATION

The importance of geometry and surface to the generation and maintenance of a stable ehd film has been discussed in the preceding sections. Lubrication is the other significant factor. Maintenance of the film demands that the ball retainer deliver to the balls in a stable manner the required amount of a lubricant with the needed properties. Stability of the film requires that the circulation of the lubricant be controlled to prevent excessive local oil buildup that can periodically cause film-thickness changes, as discussed in Section 6.1.

The demand for stability limits the total quantity of oil that can be carried in the lubrication system, but sufficient oil must be available for long life and maintenance of a low-friction coupling between the balls and retainer. Therefore, control is needed of lubricant function, quantity, and disposition. Severe demands are thus made on both the lubricant and the retainer.

4.1 Lubricant

Most precision ball-bearing gyros use oil rather than grease as the bearing lubricant. The lubricant quantity and distribution needed to assure long life is more stable in the form of oil impregnated in a porous-plastic ball retainer than grease packed around the balls.

The oil used in most gyros for more than twenty years has been Humble's Teresso V-78, a paraffinic mineral oil formulated with an anti-oxidant, an anti-foam agent, and a lubricity additive. Its nominal viscosity is 78 SSU at 210°F, or about 15 cs at 210°F. This lubricant, formulated originally as a steam-turbine oil, has performed very well in the gyro application. At various times through the years, comparative testing has been performed in attempts to find improved oil but with no marked success.

Teresso V-78 is no longer being manufactured and a replacement must soon be specified. A program to formulate and test this replacement is currently underway. The successful candidate will be one of a family of lubricants of varying viscosities for use under different operating conditions. The first approach is to match V-78 in major properties and sensitivities, thus making the substitute useful in the wide range of applications now seen by V-78. The current principal candidate, KG-80 (Kendall Refining Company), is also a paraffinic mineral oil of approximately the same viscosity as V-78. It is commercially superrefined and incorporates an anti-oxidant (Ethyl AO 702) and a boundary additive (TCP). Preliminary tests are encouraging but not yet conclusive.

There is some question as to the specific lubrication mechanism that maintains the ehd separation of the balls and races. One school presents a mechanical concept of lubrication, relating the configuration of the ehd film to bearing and environmental factors such as geometry, speed, temperature, elastic modulus, and load, and to lubricant physical properties such as viscosity and viscosity-temperature and viscosity-pressure relationships. To this concept, another school adds more chemical concerns such as composition of the lubricant, polar-component properties, surface chemical interactions, and effect of molecules adsorbed to surfaces on pressure-zone viscosity. The significance of the chemical interface of the lubricant to the metal is emphasized by the current work in the area of surface chemistry noted in Section 3.2.

Additional significant lubricant properties include thermal, oxidative, chemical and hydrolytic stability, volatility, chemical compatibility with bearing materials, and surface tension. The lubricant must, of course, be able to withstand fine filtration without detriment. It must also provide boundary lubrication under low-speed conditions. Teresso V-78 provides the properties for long successful operation; its potential replacement family hopefully will perform as well or better.

4.2 Retainer

Gyro performance, life, and torque requirements demand the use of an oil-impregnated porous-plastic ball retainer, or separator, to perform the dual functions of ball separation and provision of a lubricant reservoir and control mechanism. As demands for performance, life, wheel speed, and preload become more severe, the demands on the retainer also grow.

The most commonly used retainer material in the gyro bearing has traditionally been a paper or cloth phenolic laminate. It has had some measure of success under certain operating conditions, but has proven inadequate for the more difficult jobs. A major problem with the tubes or rods from which phenolic separators are manufactured has been lack of repeatability of physical and chemical properties from piece to piece and even along the length of a single rod. Because of the structure of the material, the retainer holds most of its lubricant on and close to the surface. Oil retention in normal phenolic-laminate retainers is only one to five percent by weight. Oil-feed characteristics are poor, and it is difficult to adjust the lubricant quantity to the narrow range between insufficient lubricant to maintain an ehd film and excess lubricant resulting in poor instrument performance.

Porous sintered nylon (Nylasint) has provided solutions to many of the problems inherent in the use of phenolic laminates. Nylasint is a through-porous material with more than twenty-five percent total porosity. Figure 22 compares laminated phenolic and Nylasint porosity characteristics. In use, the oil content is held to a value closer to fifteen percent in order to avoid the oil jag (Section 6.1) and migration problems associated with excess lubricant. The pore structure of Nylasint is bimodal, with the larger pores generally around 3.5 microns and the smaller ones around 0.6. Total porosity, pore-size distribution, sling-out characteristics, and strength are adjustable within fairly broad limits.

As a retainer material, Nylasint is not without problems. It is more difficult to machine and deburr than laminated phenolic. Its properties are better controlled than those of the phenolic, but not as well as desired, and it is weaker and softer than phenolic. It is more subject to whirl or squeal under the more rigorous operating conditions of performance, load, and speed to which it is subjected (as discussed in Section 6), though treatment of ball-pocket surfaces as well as other remedies alleviate this.

Bearing life tests and gyro performance attest to the marked superiority of Nylasint over laminated phenolic. Operating Nylasint-bearing gyros approaching 30,000 running hours are still showing excellent performance with no sign of degradation.

Both the increased oil quantity and the through-porosity of the material contribute to its success. A series of tests was conducted to establish the validity of the theory that complete circulation of the bearing lubricant occurs with use of through-porous retainers. Retainers were cut approximately in half and recemented with impermeable walls separating the two halves, as shown in Figure 23. Before rejoining, one half was impregnated with clear lubricant and the other with blue-dyed oil. Studies were made of the rate and mode of oil circulation as a function of running hours for a range of geometry, speed, lubricant, surface treatment, retainer permeability, preload, and ambient pressure. It was established that complete circulation of the lubricant does occur and that lubricant transfer takes place at the ball-to-groove and ball-to-retainer interfaces.

Dramatic evidence of the need for a through-porous retainer at higher speeds is seen in Figure 24. Typically, for a specific set of running conditions for R4 bearings, 12,000 r.p.m. life with Nylasint retainers exceeds 20,000 hours. With laminated phenolic, it ranges from 5000 to 15,000 hours; and with solid nonporous nylon with oil retention nearly equal to that of the phenolic, it approximates 2000 hours. Doubling the speed to 24,000 r.p.m. leaves Nylasint life essentially unchanged, reduces phenolic life to about 500 to 2000 hours, and drastically cuts life of solid nylon to less than 24 hours.

It is interesting to speculate on the possible role of a through-porous separator as an oil filter, since complete circulation of the oil occurs. Is particulate matter, initially in the bearing or generated on occasions of momentary asperity contact or lubricant degradation products, strained from the lubricant by the Nylasint? Dark deposits are frequently seen in the ball pockets after bearing operation.

Nylasint ball retainers perform well, but additional work is needed. Improvement in some properties, as well as in quality control, is desirable.

5. PROCESSING

Bearing processing, from completion of manufacture through gyro construction, must preserve or improve upon bearing built-in quality. In addition, this quality, as well as various performance parameters, must be monitored at critical construction stages. Development of processing techniques and evaluation means has played a major role in the evolution of gyro bearing technology.

5.1 Quality Retention

Contamination control is a critical factor in bearings that must maintain a fractional microinch ehd film stability and that do not have a frequent fresh supply of lubricant to flush out debris. Chemical, particulate, and even atmosphere contamination must be avoided.

Chemical contamination of the metal surfaces, as discussed in Section 3.2, impedes the maintenance of either boundary or ehd lubrication. Its prevention is important, particularly in final bearing manufacturing and early instrument-processing stages. Exposure to various cleaning agents and processing and storage fluids must be evaluated for potential deposition of unwanted films or detrimental surface chemical reactions. Housekeeping practices must be well controlled to prevent contaminated fluids, dirty glassware, improper processing, or human contamination from affecting the metal surfaces. At the moment, the only reasonably economical nondestructive test for monitoring a subtle form of chemical contamination is the spreading rate of oil on the surface.

Particulate contamination is a more commonly recognized problem, and at one point abrasive contamination during lapping was the cause for many early bearing failures. Recognition of this problem by microscopic inspection with polarized light led to improved cleaning techniques, and this issue has been largely resolved.

There are many potential sources of particulate contamination, as in bearing manufacture, packaging, instrument construction, in the solvents or the lubricant, and from the ball retainer. Soft as well as hard particles can be detrimental under the high pressures of the films between the balls and races. Ball-retainer deburring and cleaning are particularly significant in the prevention of particulate contamination, especially since the retainer remains a possible source of contamination throughout the running life of the bearing. Figures 25 and 26 are examples of brinelling by soft and hard particles.

Atmosphere contamination refers to the condition of the ambient atmosphere in which the bearings are run, both in test chambers and in the final-sealed gyro float. One concern is control of the atmosphere to prevent detrimental chemical changes in the lubricant or on the metal surface. The other is retention of the physical properties needed for acceptable windage torque, heat transfer, and float leak detection.

Another area of concern in bearing processing is retention of bearing geometry. Factors associated with bearing geometry, potentially affected by variations due to processing, include basic instrument design sensitivities as well as bearing dynamics and ehd film generation and maintenance. Of obvious concern are the changes in race-groove roundness that occur with interference fitting of out-of-round shafts or wheels to round bearing bores and outside diameters; a representative example is shown in Figure 27. Less obvious, but equally significant, are out-of-square and out-of-parallel clamping distortions.

A gross geometric change within the race grooves, attributable to processing, is brinelling due to overloading, as shown in Figure 28. This can occur during faulty preload application, inadvertent overload due to fixturing or assembly problems, or just poor handling. It presents a serious problem, and its detection is very important to prevent further processing of bearings that will fail later due to this damage.

5.2 In-Process Testing

The extreme importance and inaccessibility of the bearing package in the gyro make it mandatory that in-process quality-assurance tests be conducted. These tests must detect conditions that might lead to early bearing failure or poor instrument performance. Various test methods and devices have been developed for this in-process evaluation.

When the bearing is accessible in the unassembled condition, as in early processing stages or in diagnostic testing following problem detection, a wide range of test methods is available. These include microscopic inspection with normal or polarized light and with the interference microscope, which is particularly useful for recognizing and characterizing topographic aberrations. Also available are various means for measuring and tracing geometry and surface finish. In addition, a wide range of destructive tests can be performed for diagnosis or quality control.

One of the most useful test devices for assembled bearing surface and lubricant characterization is the low-speed dynamometer (LSD), illustrated along with representative traces in Figure 29. It consists of a spindle on which is mounted the bearing outer race (or races in the case of a preloaded bearing pair), a dead-weight axial-loading system for a single test bearing, a means for rotating the spindle slowly (generally at 1 rpm), a beam on which a strain gauge is mounted to restrain the inner race or races from rotation, and a means for recording the strain-gauge output. Provision is also made for reversal of direction of rotation, load and speed variation, zero setting, and calibration. The resultant torque trace tells a great deal about the torque level, metal surfaces, lubricant condition, contamination, and geometry. For example, a high hash level generally characterizes a poor surface finish, contamination, or lubricant or metal degradation, the specifics of which are readily determinable by other means. Individual trace features show metal damage or dirt; the former is characterized by an initial sharp torque drop followed by a rise, whereas the dirt shows an initial increase. Spacing of torque disturbances pinpoints discontinuity location as being inner race, outer race, or ball.

The LSD can be used following various critical processing stages of the individual bearing or preloaded pair until the wheel package is sealed into the float. Changes in trace character rather than absolute levels are monitored as bearing degradation symptoms. One of the most valuable functions of the LSD is early detection of lubricant degradation. It frequently provides the first test to disclose bearing deterioration problems and is therefore very important. Perhaps just as important is the user's willingness to conduct the test and act upon the evidence.

The other of the two most important diagnostic tools is the recording milliwattmeter-dynamometer, generally called the wattmeter or milliwattmeter. The milliwattmeter is used at bearing operating speed and simply provides a sensitive (zero-suppressed) trace of motor-power input. Assuming a stable power supply, variations in the power trace reflect variations in the bearing torque demand, which in turn generally correlate with variations in factors influencing gyro wheel axial position. Thus the wattmeter provides a measure of potential gyro performance. Figure 30 shows milliwattmeter traces representing various classes of gyro performance.

A particularly valuable aspect of the wattmeter is its usefulness in a wide range of test configurations, including bearings assembled in a test fixture, final wheel and bearing assembly, sealed float, and the completed gyro. Thus it is a useful diagnostic and research tool as well as a valued in-process tester. Some of the specific behavior patterns monitored by the wattmeter are discussed in Section 6.

Other test devices fulfill some of the same functions as the milliwattmeter. One is a high-speed dynamometer which has a torque readout, and therefore does not depend upon power-supply stability. Its usefulness extends to the float assembly stage. Various other high-speed torque testers are in use both as research and production tools.

Another useful series of in-process tests is a group using wheel runup and rundown for bearing and motor torque evaluation. Deceleration at high speed yields bearing friction and windage torque data, while the low-speed end is essentially not influenced by windage torque. Total rundown time provides a rough monitor of stability of running conditions. These tests are useful over a wide range of gyro construction steps and are particularly useful in gyros subjected to prolonged or repeated testing. Bearing degradation can be detected by this test as well as by milliwattmeter evaluation, but generally not until hundreds of running hours after the inception of failure.

Various other test methods are useful either as in-process steps or as special diagnostic tests, depending upon the requirements of the particular program. One of these is the mounting of a wheel package or float on a cradle with vibration pickups and monitoring the bearing dynamics at the retainer and ball-group frequency. Some bearing configurations demonstrate performance particularly correlatable with ball-group frequency.

Another useful test device is the inner-package evaluator. This is a temperature-controlled hydrostatic-gas-bearing-supported horizontal element restrained by a torque feedback loop on which a bearing package or float can be mounted and the wheel run. Recordings are taken of torque to balance, ball-group-frequency vibration amplitude about the output axis, and motor-power input. Though this device is far less sensitive and less versatile in discerning disturbing inputs than a completed gyro, it is very useful.

Many in-process test methods are available for bearing package evaluation, as shown in Figure 31. Their usefulness is a function of the specific instrument requirements and problems. The two most useful functional test devices through the years, though, have been the low-speed dynamometer and the recording milliwattmeter-dynamometer.

6. PERFORMANCE

Gyro performance reflects, among other factors, spin-axis bearing package performance. Some aspects of bearing dynamic behavior affect instrument precision without influencing bearing life, but bearing degradation almost always causes gyro performance degradation.

6.1 Bearing Dynamics

Though many aspects of bearing dynamic behavior are now understood, there are still gaps in the understanding of the fundamentals. The old theories, based on a model with metallic contact between race and ball, do not adequately account for viscous effects or for ball-retainer coupling. Work is currently underway to close these gaps.

One of the earliest of the bearing dynamics phenomena to be explained was the "classical jag", or "oil jag". The jag was first seen in gyro performance as an abrupt change in float balance along the spin axis, followed by an exponential decay. Later development of the milliwattmeter disclosed the existence of a simultaneous sharp increase in motor power, also followed by an exponential return. Microscope observation of stroboscopically illuminated bearings showed the power increase to be accompanied by the centrifugal release of an oil droplet from the ball retainer OD to the outer-race groove. This is a brief version of an investigation that spanned several years in various places by various people. It is now apparent that oil is deposited by the balls in the retainer ball pockets and that it runs out and tends to collect on the retainer OD . When the centrifugal force on the oil drop exceeds that of the surface tension, the drop is thrown to the outer race. If it lands in the race groove, it presents the balls with a sharp increase in film thickness. This film thickening in one bearing increases axial load on both bearings and moves the wheel center of gravity, thus changing the gyro float balance. Bearing torque and therefore motor power also increases, due to added viscous drag and increased load. Equilibrium conditions return exponentially.

As expected, improved control of lubricant quantity reduced the severity of the problem. Bearing design also influences jagging incidence and severity. For example, the problem becomes more acute with a full outer race than with one with a land ground off, and is alleviated by a more open groove-to-ball conformity.

Another aspect of bearing dynamics that has yielded to investigation is the interaction of the ball groups in a preloaded bearing pair at their beat frequency. A characteristic sinusoidal milliwattmeter trace reflects the varying bearing torque as the two misaligned ball groups beat with respect to each other, and the effective preload is increased and decreased by the varying phase relationships of the larger balls in the two ball groups. In addition, each ball varies its speed and moves across its ball pocket (at the beat frequency), and the ball retainer center of gravity is driven in a circular path around the bearing axis (also at the beat frequency) with respect to the ball group. Modification of this dynamic behavior can be achieved by varying the size match of the balls and their relative position in each bearing, and by varying basic ball-group size of one bearing with respect to the other, thus changing the contact-angle match and therefore the beat frequency.

Variation of the beat frequency occurs as the direction of an acceleration (e.g. gravity) is changed with respect to the spin axis. The wheel mass acted upon by the changing acceleration direction alters the effective load on the two bearings differentially, and therefore affects their contact angles and beat frequency.

High-speed ball-retainer whirl (eccentric motion of the retainer center of gravity around the bearing axis), and in a limiting case squeal, are serious bearing dynamics problems. Synchronous whirl, or translation of the retainer center of gravity around the bearing axis at the wheel frequency, is not generally a problem. Its severity is a function of wheel unbalance and bearing preload, generally occurring only in lightly loaded bearings. More serious is high-frequency whirl caused by ball-to-retainer frictional coupling, which, in generation mechanism, is similar to journal-bearing dry-friction whirl. Severe cases of this whirl are accompanied by very high erratic torque and in many cases by audible squealing or chirping. This condition appears on the milliwattmeter as a power disturbance with a high-erratic, hashy trace. It may be momentary or intermittent, appearing as a sharp short power spike, undetectable by other means. The squeal condition is intolerable and, depending upon its severity, it is accompanied by poor gyro performance, excessive torque, and early bearing failure.

The best method of correction is reduction of the frictional coupling of the ball to retainer, generally by improved lubrication. In the case of the Nylasint retainer, this can be accomplished by salt-blasting the ball pockets, which compacts the surfaces and provides better lubrication. Other means for compacting the ball-pocket surfaces are also effective.

Another approach is to randomize the driving mechanism of retainer squeal. This can be done by deliberately mismatching ball sizes within the bearing or by spacing the ball pockets nonuniformly in the retainer. Though effective, these methods exact a penalty in higher bearing torque.

Additional investigations into the fundamentals of bearing dynamics are in progress. One test device rotates the inner and outer races of an axially loaded bearing in opposite directions, as seen in Figure 32, maintaining the ball group and retainer (when one is used) fixed in space. This permits microscope observation of ball motions and use of instrumentation to measure retainer forces. Some tests are conducted with balls with small diametral through-holes, providing a preferred ball-rotation axis and an observation and analysis tool for ball-motion study. Another device permits outer-race rotation of a preloaded bearing pair, with independent torque measurements of the two bearings. It is also instrumented for axial and radial retainer-motion monitoring.

Studies conducted with these test devices will disclose fundamental bearing knowledge. Factors being studied include ball precession, ball slip as shown in Figure 33, retainer forces, ball-group speed ratio, retainer dynamics, and bearing torque versus variables of speed, lubricant type and quantity, geometry, load, retainer configuration, metal surface characteristics, and others. These studies will lead to continued improvement in gyro life and performance.

6.2 Gyro Performance

The gyro itself is the only device sensitive enough to tell whether or not bearing performance goals are achieved. In addition, it is a very useful diagnostic tool for bearing parameter evaluation and improvement. Finally, it can be easily monitored to disclose the health of the bearing package at any time during the life of the gyro.

Gyro stability, determined by a variety of tests in both fixed and varying acceleration fields, combines factors associated with and independent of the bearing package. Various means are used to separate bearing performance from other factors. One common test, for example, uses the gyro to stabilize a servo-controlled turntable in inertial space, thus causing the gyro to tumble in the earth's gravity field. Repeatability of gyro unbalance from revolution to revolution provides a record of balance stability. Comparison of this stability for gravity positions in which random motion along the spin axis will show as torque uncertainties, with positions not affected by spin-axis instability, provides an indication of bearing performance. Coupling of this information with simultaneous recordings of motor input power and signal-generator output at the ball-group frequency provides a more complete picture, as shown in Figure 34.

Various other tests of float balance with the gyro tumbling and in various fixed-gravity orientations, and subjected to higher vibratory and steady-state acceleration levels, reveal bearing package data. In each case, the power and output frequency analyses are necessary correlative tools. In fact, these tests, combined with specific bearing parameter or running condition variables, provide useful design information.

No ball-bearing gyro should experience severe bearing failure without warning. As bearings degrade, they show very distinctive symptoms. Perhaps the first sign of impending bearing failure is evidenced by gradually degraded performance, particularly with regard to spin-axis stability. The milliwattmeter trace develops erratic periods as degradation progresses, with both degree of power variation and ratio of rough-to-smooth trace becoming greater. Changes may also occur in the signal at the ball-group frequency. With continuing degradation, wheel deceleration tests show changes in both character and total rundown time. The time from the first detectable failure symptoms until appreciably higher power demand occurs can amount to several hundred or thousands of running hours.

Use of current ball-bearing technology has permitted the construction of gyros demonstrating performance levels unexcelled by any other spin-axis suspension. Long life has also characterized these instruments, both in terms of wheel running hours and shelf life, examples of which are cited in Table II. Successful bearing use does demand, however, rigorous adherence to the highest quality control standards in both manufacture and processing. Some of these standards have been discussed in this paper.

It is expected that future efforts in ball-bearing technology will yield further advances in both understanding of fundamentals and development of improvements. Work is being performed, for example, in the fields of bearing dynamics, the lubrication mechanism, and the race-groove and ball surfaces. These efforts should result in more consistent achievement at lower cost of currently demonstrated excellent performance and long reliable life. They should also provide the basis for meeting even more rigid requirements in the future.

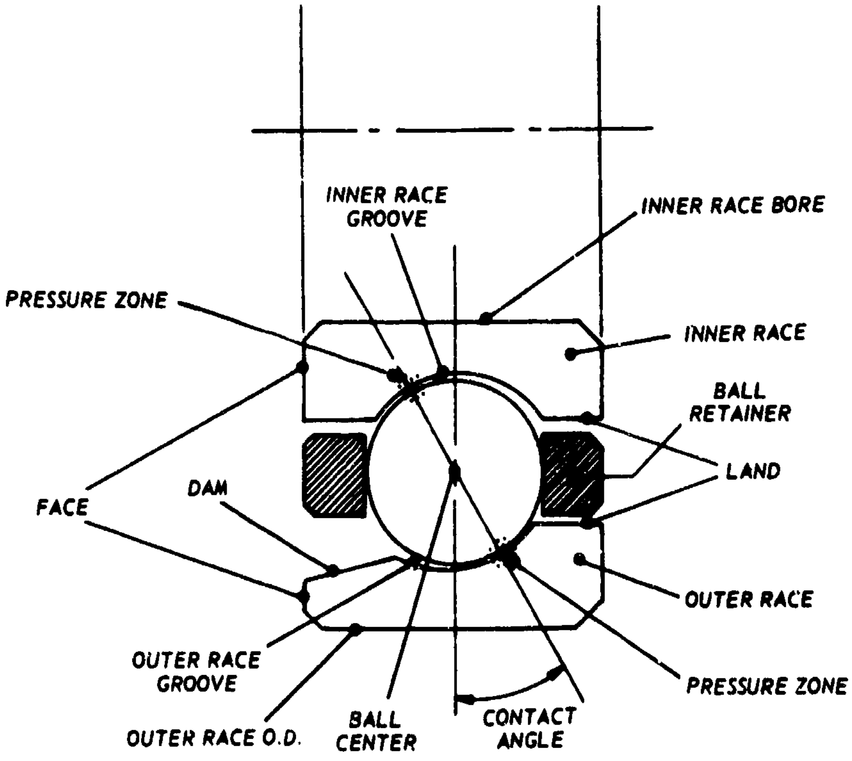
ACKNOWLEDGMENTS

This paper covers work on gyro ball-bearing technology carried out under various US Air Force, US Navy, and National Aeronautics and Space Administration contracts. Current work in the bearing dynamics area is being performed under the auspices of DSR Project 52-306, sponsored by the Avionics Laboratory of the Air Force Systems Command through Contract F33615-68-C-1155 with the Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts. Current work in the area of surface chemistry is being carried out under the auspices of DSR Project 55-219, sponsored by the Manned Spacecraft Center of NASA through Contract NAS 9-3079 with the MIT Instrumentation Laboratory. In the lubricant replacement program, current effort is under the auspices of DSR Project 52-275, sponsored by the Space and Missile Systems Organization of the Air Force Systems Command under contract AF 04(694)-999 with the MIT Instrumentation Laboratory.

The contributions of various governmental and private agencies to the status of the bearing technology reported here are also acknowledged.

APPENDIX

Bearing Nomenclature



$$\text{RACE GROOVE CURVATURE} = \frac{\text{RACE GROOVE RADIUS}}{\text{BALL DIAMETER}} \times 100\%$$

BIBLIOGRAPHY

1. *Gyro Spin-Axis Hydrodynamic Bearing Symposium.* Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 12-14 December 1966. References 2-15 below are from the Ball Bearing Proceedings.

2. Denhard, W.G. *Cost Versus Value of Ball Bearings.* Report E-1990, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, July 1966.

3. Lement, B.S. et al. *Quality of Steels Used in Gyro Ball Bearings.* Man Labs Inc., Cambridge, Massachusetts, December 1966.

4. Palmieri, J.R.; Allen, S. *High Treatment of Bearing Steels.* Report E-2085, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, December 1966.

5. Farago, F.T. *Dimensional Measurements of Gyro Bearing Components by Autocollimation and Other Advanced Methods.* New Departure-Hyatt Bearings, Division of General Motors Corporation, 1966.

6. Orcutt, F.K.; Cheng, H.S. *Lubrication of Rolling-Contact Instrument Bearings.* Mechanical Technology, Inc., Latham, New York, December 1966.

7. Allen, S.; Palmieri, J.R. *A Metallurgical Modification Caused by Finish Operations on Ball-Bearing Race Grooves.* Report E-2084, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, December 1966.

8. Gereg, C.V. *Lubricant Film Electrical Resistance Gage - Its Application to Ball Bearing Raceways.* The Barden Corporation, Danbury, Connecticut, 1966.

9. Hannan, C.H. *Observations of Oil and Water Absorption Characteristics of Spin Axis Bearing Retainers.* Miniature Precision Bearings, Inc., Keene, New Hampshire, 1966.

10. Leveille, A.R. *Properties of a Power Function Model of Ball Bearing Axial Load versus Deflection.* The Barden Corporation, Danbury, Connecticut, 1966.

11. Roberts, M. *A Study of Oil Circulation in the R4 Spin-Axis Bearing with Sintered Nylon Ball Retainer.* Report E-2082, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, December 1966.

12. Klaus, E.E. *A Study of Deposit Forming Tendencies with Gyro-Bearing Lubricants.* Petroleum Refining Laboratory Division, Chemical Engineering Department, The Pennsylvania State University, University Park, Pennsylvania, 1966.

13. Singleterry, C.R. *Some Factors Affecting the Movement of Oil over Metal Surfaces.* Naval Research Laboratory, Washington, DC, 1966.

14. Wyatt, O.H. *British Experience with Gyro Spin Axis Ball Bearings with Special Reference to Geometric Factors.* Royal Aircraft Establishment, Farnborough, Hampshire, England, 1966.

15. Kingsbury, E.P. *Torque Variations in Instrument Ball Bearings.* Report E-1675, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1964.

16. *Bearings Conference.* Dartmouth College, Hanover, New Hampshire, 7-9 September 1966. References 17-28 below are from the Proceedings.

17. Draper, C.S. *Bearings.* Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1966.

18. Griffths, C.A. *Retainer Instability in Thin Section Instrument Bearings.* Split Ball Bearing Division, Miniature Precision Bearings, Inc., Lebanon, New Hampshire, August 1966.

19. Grant, C.L. et al. *Measurement of Oil Movement in Ball Bearings with Porous Retainers.* University of New Hampshire, Durham, New Hampshire, 1966.

20. Dane, E.B. *The Manufacture of Consistently Good Ball Bearings.* Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, June 1966.

21. Devine, M.J. et al. *Anti-Friction Bearing Concepts for Non-Conventional Lubricants.* Naval Air Materials Laboratory, Philadelphia, Pennsylvania, 1966.

22. Esterly, D.D. *Using Bearing Characteristic Frequency Amplitudes to Predict Bearing Performance.* Ampex Corporation, Redwood City, California, July 1966.

23. Zaretsky, E.V. Anderson, W.J. *Material Properties and Processing Variables and Their Effect on Rolling-Element Fatigue.* NASA TMX-52227, Lewis Research Center, Cleveland, Ohio, 1966.

24. Ehrich, F.F. *Bearing Influence on Rotor Dynamics.* General Electric Company, West Lynn, Massachusetts, August 1966.

25. Given, P.S. *Lubricant Film Effects on Rolling Contact Fatigue.* SKF Industries Inc., Philadelphia, Pennsylvania, September 1966.

26. Cron, R. *Lubrication to Improve Reliability of Components of Servomechanisms.* US Naval Ammunition Depot, Crane, Indiana, 1966.

27. Jones, A.B. Jr Poplawski, J. *A Computer Study of the Effects of the Various Design Parameters and Operating Conditions on Rolling Element Bearing Performance.* Pratt & Whitney Aircraft Corporation, Hartford, Connecticut, 1966.

28. Hannan, C.H. *Chemical Treatment of Rolling Element Bearing Surfaces.* Miniature Precision Bearings Inc., Keene, New Hampshire, 1966.

29. *Elastohydrodynamic Lubrication.* Symposium conducted by the Lubrication and Wear Group of the Institution of Mechanical Engineers, Leeds, Proceedings 1965-66, Vol.180, Part 3B.

30. *Contact Angle, Wettability, and Adhesion.* The Kendall Awards Symposium honoring W.A.Zisman, 2-3 April 1963. Advances in Chemistry Series No.43, American Chemical Society, Washington, DC, 1964.

31. Bowden, F.P. / Tabor, D. — *The Friction and Lubrication of Solids.* Parts I and II. Oxford University Press, London, 1954 and 1964 respectively.

32. Cameron, A. — *The Principles of Lubrication.* John Wiley, 1966.

33. Denhard, W.G. et al. — *Failure Analysis of Critical Ball Bearings.* Report E-1781, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, April 1965.

34. Dowson, D. / Higginson, G.R. — *A Numerical Solution to the Elastohydrodynamic Problem.* Journal of Mechanical Engineering Science, Vol.I, 1, 1959.

35. Essex. A.R. — *Measurement and Generation of Truly Circular Form.* Paper No.15 of Gyro Symposium, Institution of Mechanical Engineers, London, 1965.

36. Fein, R.S. — *Possible Role of Compressional Viscoelasticity in Concentrated Contact Lubrication.* ASME Transactions, 89F, 1967.

37. *Final Report on the Gyro Spin-Axis Bearing Program.* Report R-418, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, September 1963.

38. Fitzsimmons, V.G. et al. — *A New Approach to Lubricating Ball Bearings.* NRL Report 6356, US Naval Research Laboratory, Washington, DC, 28 December 1965.

39. Furey, M.J. — *Surface Roughness Effects on Metallic Contact and Friction.* American Society of Lubrication Engineers Transactions, Vol.6, 1963.

40. Harris, T.A. — *Rolling Bearing Analysis.* John Wiley, 1966.

41. Hersey, M.D. — *Theory and Research in Lubrication.* John Wiley, 1966.

42. Horsch, J.D. — *Correlation of Gyro Spin-Axis Ball Bearing Performance with the Dynamic Lubricating Film.* American Society of Lubrication Engineers Transcations, Vol.6, No.2, 1963.

43. Howles, D.J. / Munson, H.E. — *Summary Report, Spin-Axis Bearings.* N66-35016, NASA Scientific and Technical Information Facility, 27 June 1966.

44. Jones, A.B. — *Analysis of Stresses and Deflections.* New Departure, 1946.

45. Kannel, J. et al. — *A Study of the Influence of Lubricants on High-Speed Rolling-Contact Bearing Performance.* ASD-TDR-61-643, Parts VI and VII, August 1966 and December 1967 respectively.

46. Palmer, P.J. — *Precision Inertial Gyro Testing at MIT.* AGARD Lecture Series No.XXX on Inertial Component Testing - Philosophy and Methods. Paris, 24-29 June 1968.

47. *Pressure-Viscosity Report.* Vols.I and II, American Society of Mechanical Engineers Research Publication, 1953.

48. Tallian, T.E. et al. — *Lubricant Films in Rolling Contact of Rough Surfaces.* American Society of Lubrication Engineers Transactions, Vol.7, No.2, April 1964.

TABLE I

Effects of TCP Coating on Low-Speed Endurance and Torque

R4 bearing, 1 r.p.m. 10 lb axial load (4.5 kg)	*As finished*	*TCP coated*
Torque (gm cm)	2 to 6	2 to 3
Time to failure (hours)	1 to 5	200 to 1000+

TABLE II

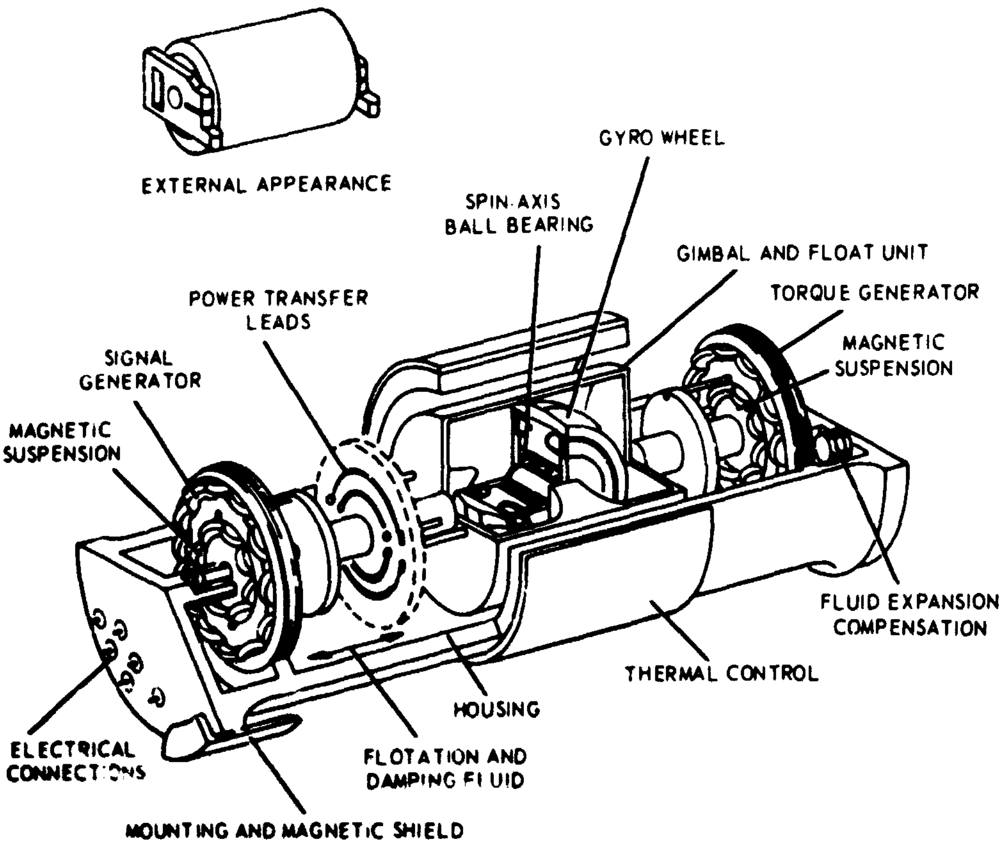
Gyro Bearing Running Time[*]

System	*Gyro*	*Total running time*		*Gyro age at most recent bearing operation (years)*
		hours	*years*	
A	6-1	28,700	3.3	5.7
	6-2	28,300	3.3	5.7
	6-3	26,500	3.0	5.4
B	6-4	26,500	3.0	4.0
	6-5	25,200	2.8	3.4
	6-6	25,600	2.9	3.7
	6-7	22,500	2.6	3.9
C	9-1	13,900	1.6	2.9
	9-2	8,900	1.0	2.6
	9-3	6,400	0.8	1.4
D	2-2	5,300	0.6	7.7
	2-3	8,500	1.0	7.4
	2-4	6,700	0.8	7.3
Gyro Life	1-1	31,600[†]	3.6	6.0
Tests	2-1	48,300[**]	5.5	8.5

[*] Data as of 1 January 1968

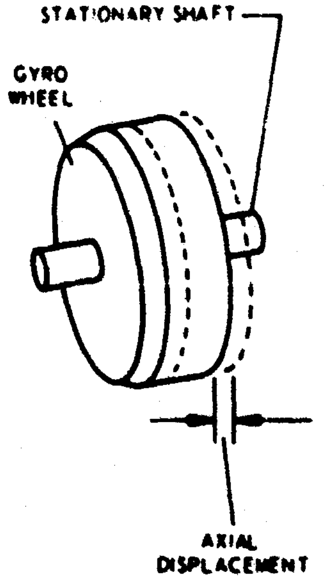
[†] Performance failure at ≃ 5000 hours (0.6 years)

[**] Still running



PICTORIAL SCHEMATIC

Fig. 1 Single-degree-of-freedom floated gyro



FOR ONE MILE PER HOUR NAVIGATIONAL ACCURACY ON EARTH'S SURFACE (A gyro drift rate of < 1.0 meru, or 0.015°/hour)	
THE WHEEL LOCATION STABILITY REQUIRED IS 0.2×10^{-6} INCH OR 50 ANGSTROMS	THE MAXIMUM OIL DROP MIGRATION PERMITTED ACROSS A SINGLE (R4) BEARING IS 0.3 MILLIGRAM OR 0.030 INCH DIAMETER
GYRO WHEEL: MASS – 225 GRAMS ANGULAR MOMENTUM – 2×10^{6} GM-CM^2 SEC	

Fig. 2 Navigational accuracy versus stability requirements at one g

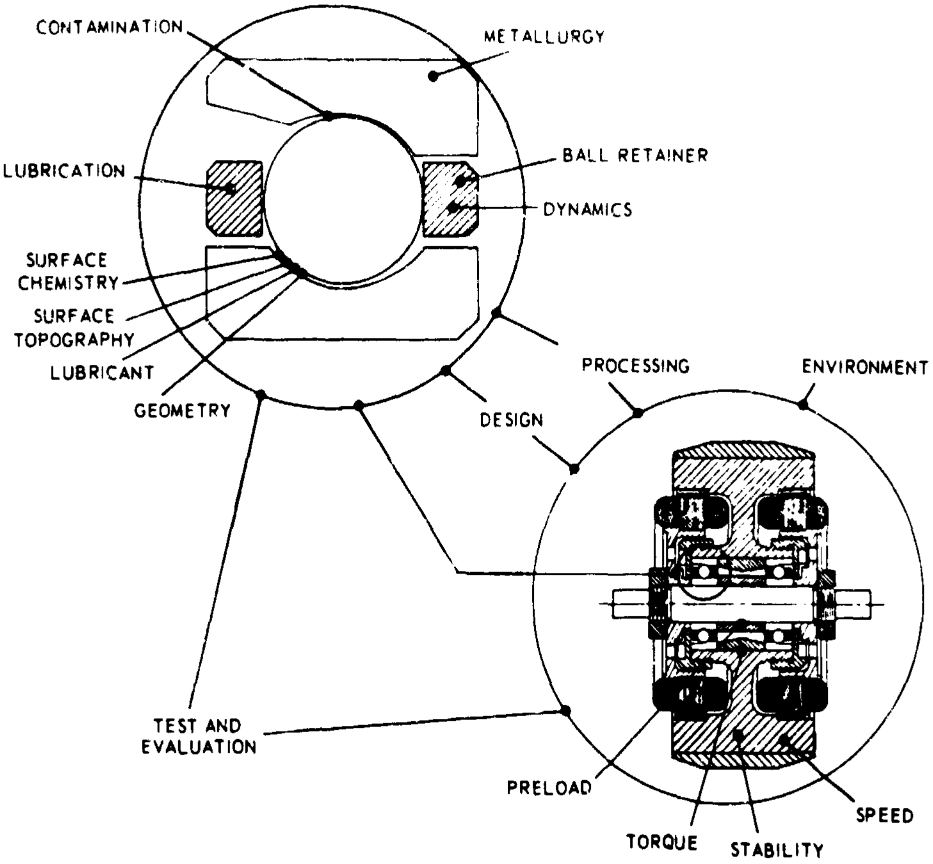


Fig.3 Major areas of bearing progress

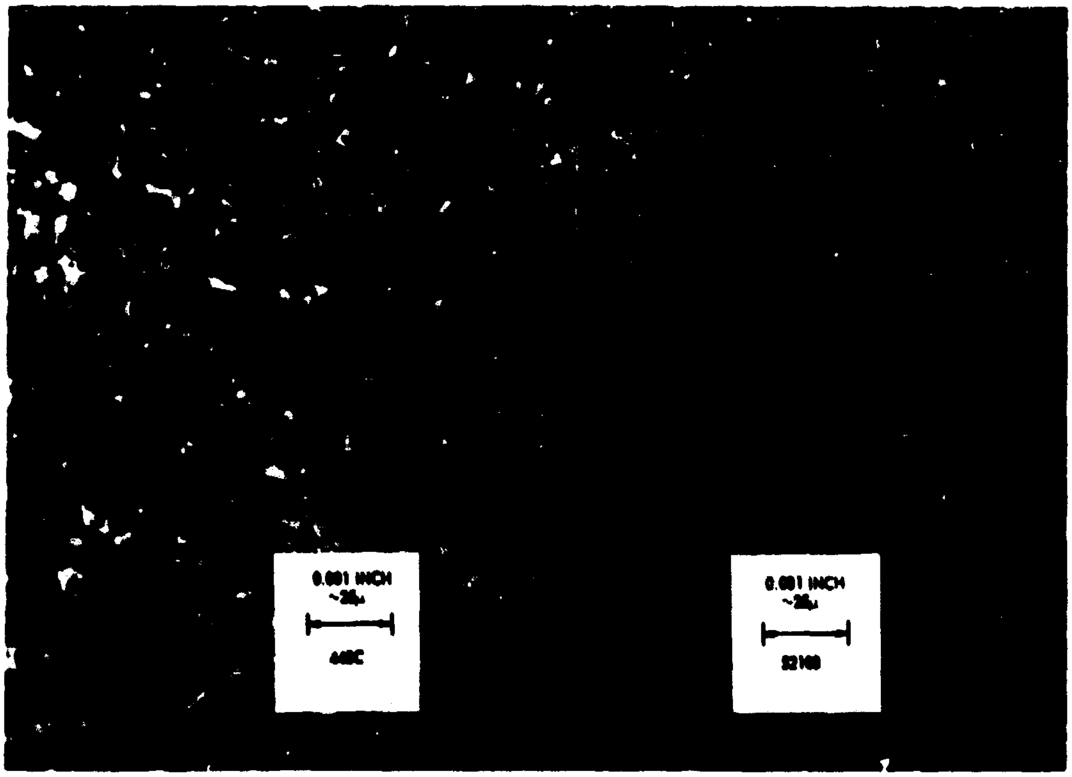
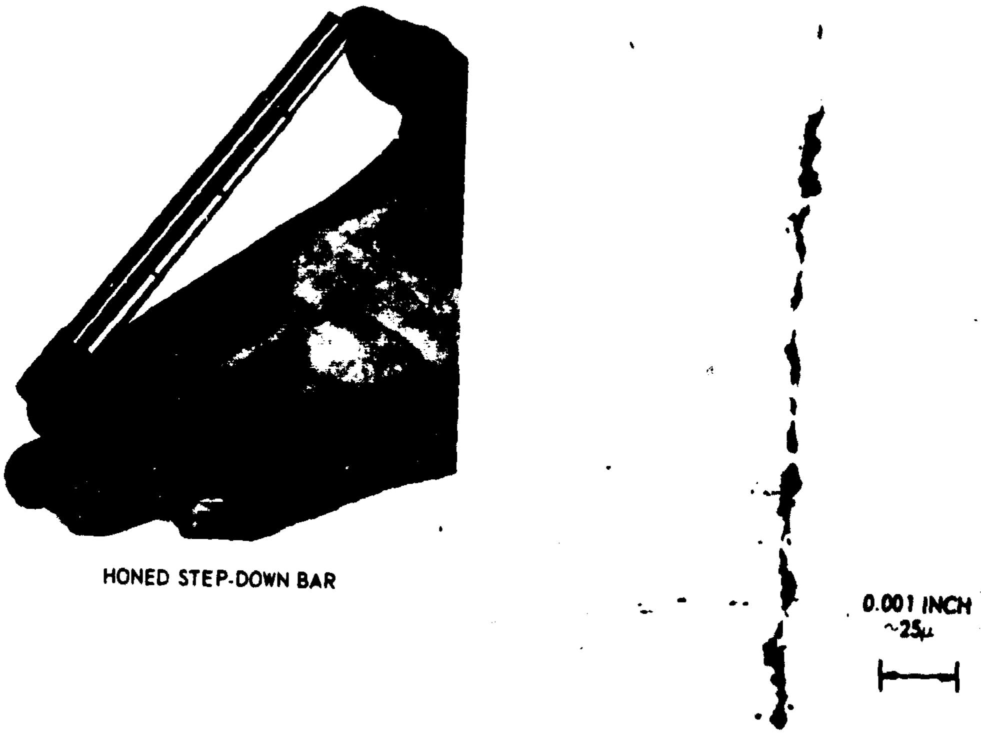


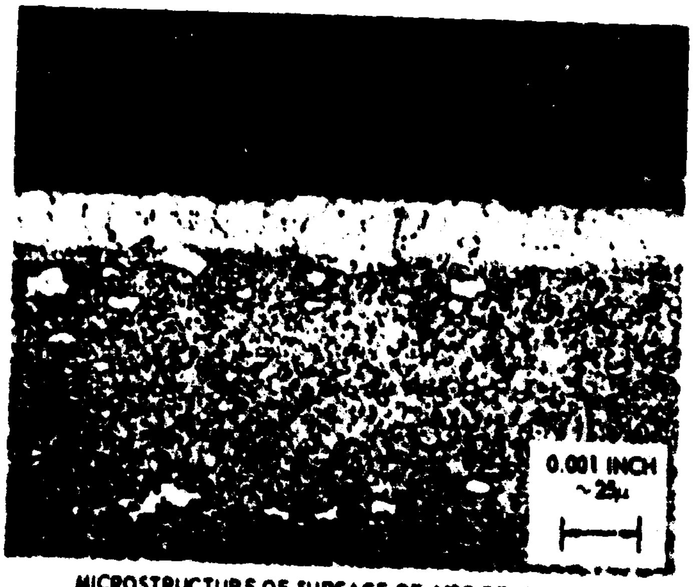
Fig.4 Comparison of microstructure of 440C versus 52100 steel



HONED STEP-DOWN BAR

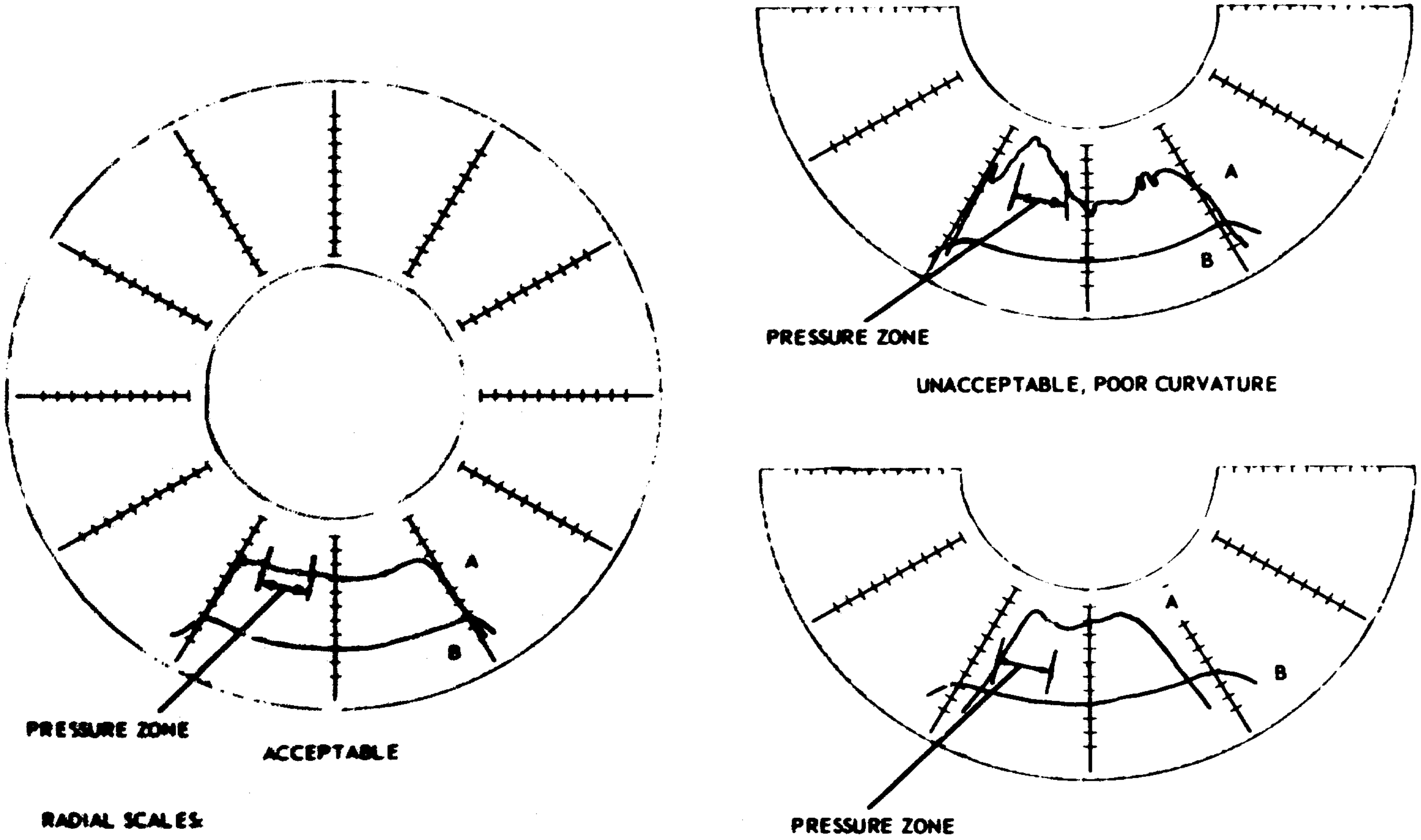
STRINGER INCLUSION IN BAR

Fig.5 Inspection of metal for inclusions



MICROSTRUCTURE OF SURFACE OF 440C BEARING SHOWING DECARBURIZATION

Fig.6 Heat-treatment surface modification



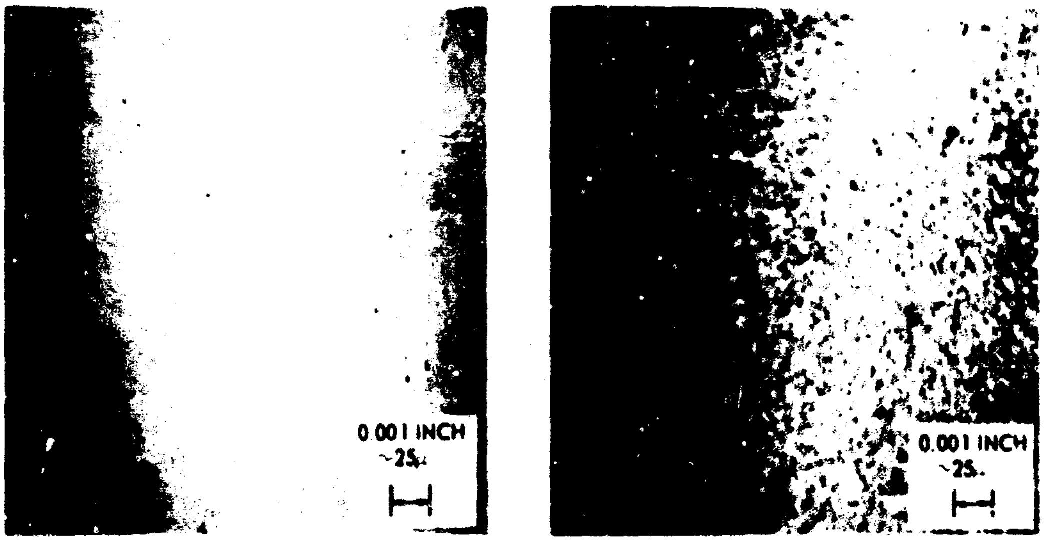
ACCEPTABLE

UNACCEPTABLE, POOR CURVATURE

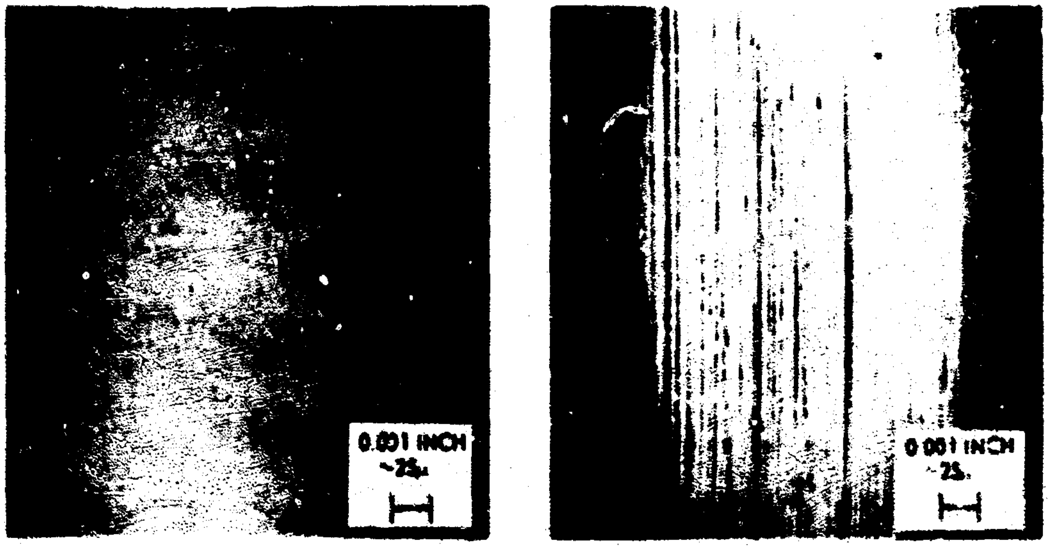
UNACCEPTABLE, EXCESSIVE EDGE ROUNDING

RADIAL SCALES:
A TRACE = 10 MICROINCHES/DIV (~0.25μ/div)
B TRACE = 500 MICROINCHES/DIV (~12.5μ/div)

Fig.7 Bearing race-groove cross-curvature Talyrond traces



ETHYLENE GLYCOL PRERUN BALL LAP



DOUBLE HONE STRING LAP

Fig.8 Acceptable race-groove surface finishes obtained by various techniques

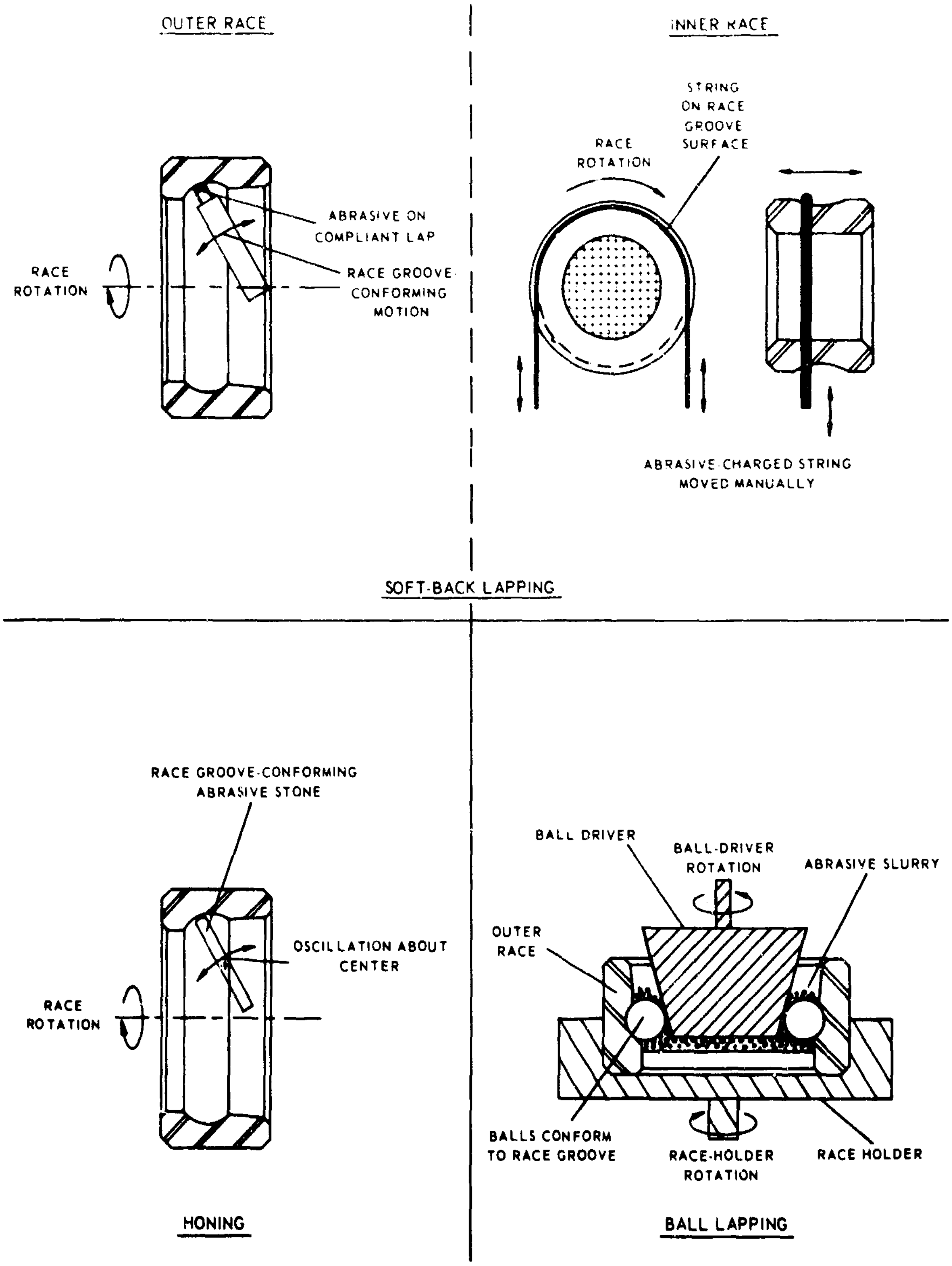
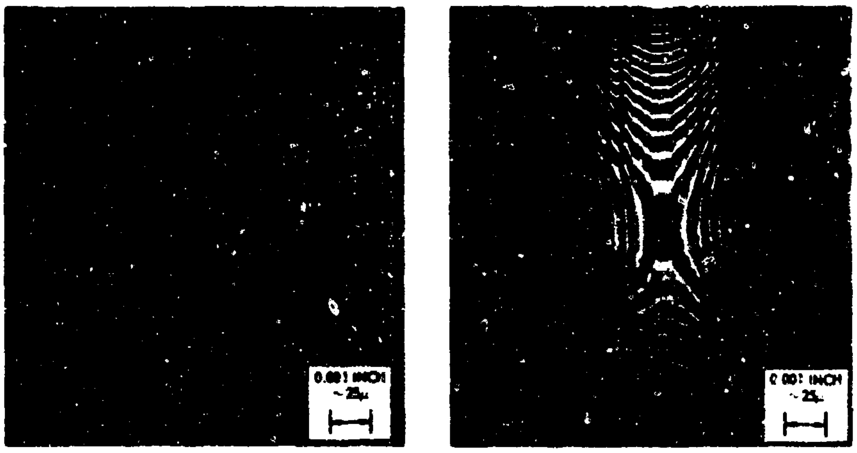


Fig. 9 Race-groove surface-finishing techniques



NORMAL MICROSCOPIC APPEARANCE

INTERFERENCE FRINGE PATTERN OF SAME FIELD SHOWING COMET IS RAISED. FRINGE SPACING IS 10.6 MICROINCHES (0.27μ).

Fig. 10 Comet on bearing inner race-groove surface

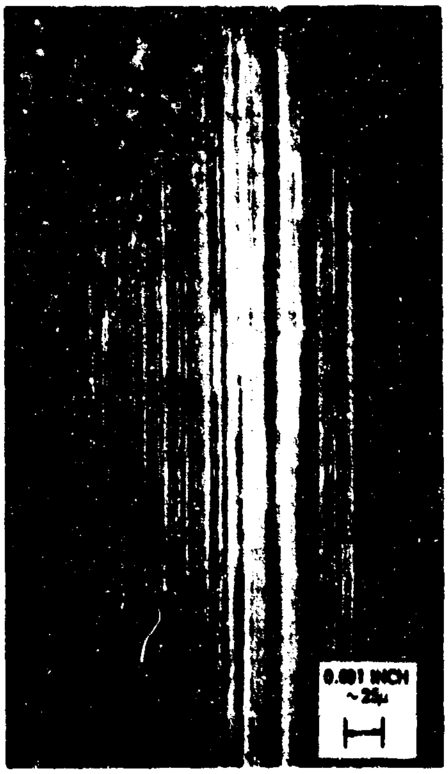
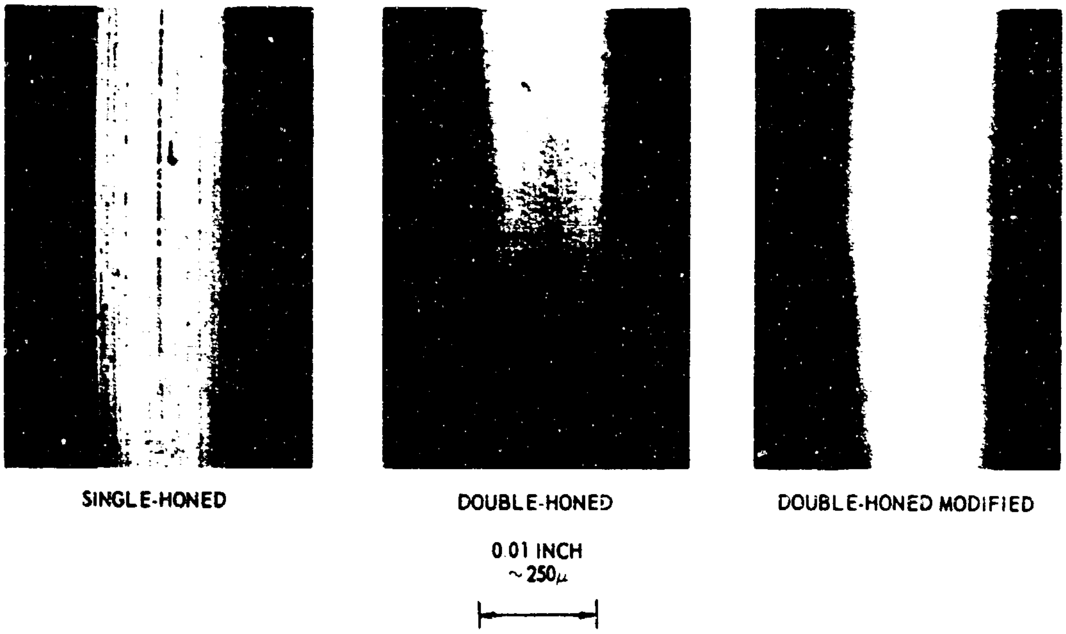


Fig. 11 R4 bearing outer-race groove after 10,000 hours of successful operation at 24,000 r.p.m.

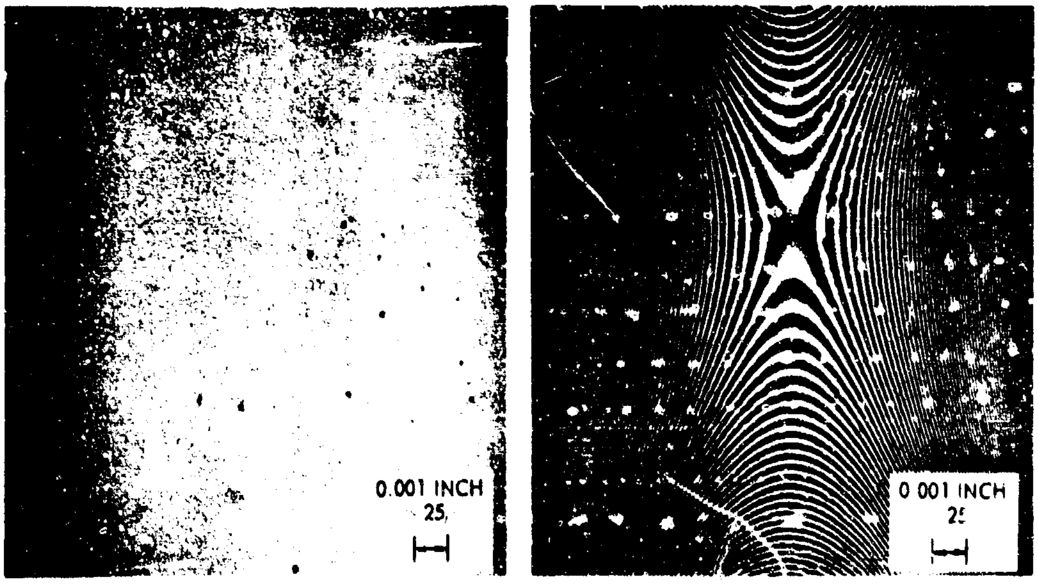


SINGLE-HONED

DOUBLE-HONED

DOUBLE-HONED MODIFIED

Fig. 12 Progress in honed 440C outer race-groove finishes



NORMAL MICROSCOPIC APPEARANCE

INTERFERENCE FRINGE PATTERN OF SAME FIELD FRINGE SPACING IS 10.6 MICROINCHES (0.27μ).

Fig. 13 Typical ball-lapped finish, R4 440C inner-race groove

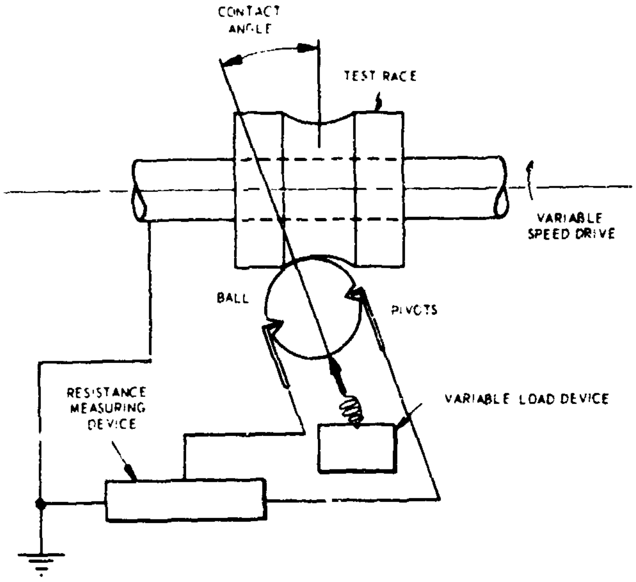
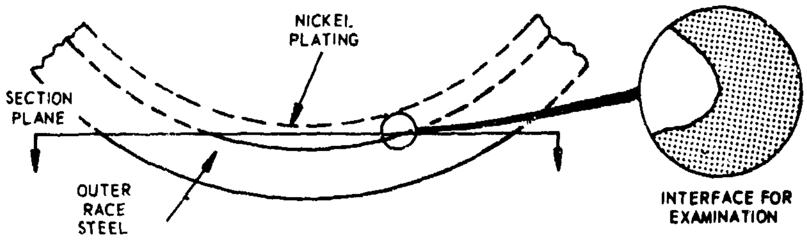
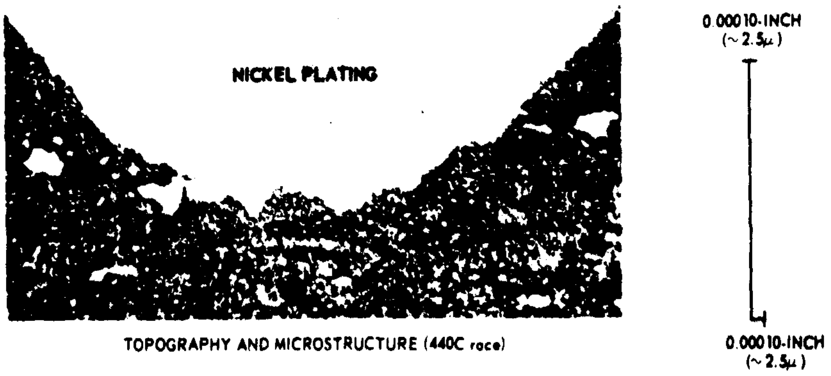


Fig. 14 Lubricant-film electrical-resistance gauge



SECTIONING METHOD FOR OUTER RACE GROOVE



TOPOGRAPHY AND MICROSTRUCTURE (440C race)

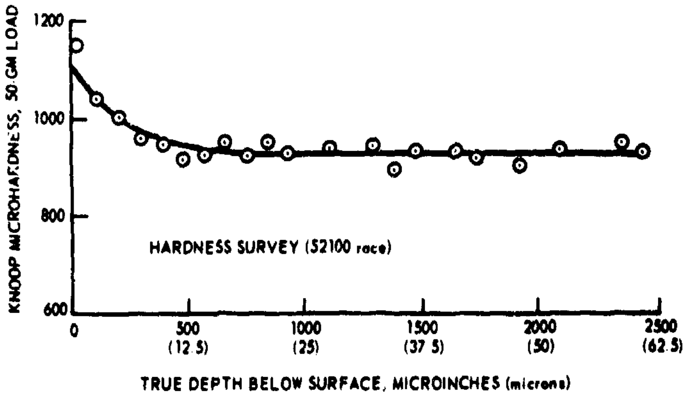


Fig. 15 Taper sections of outer-race grooves

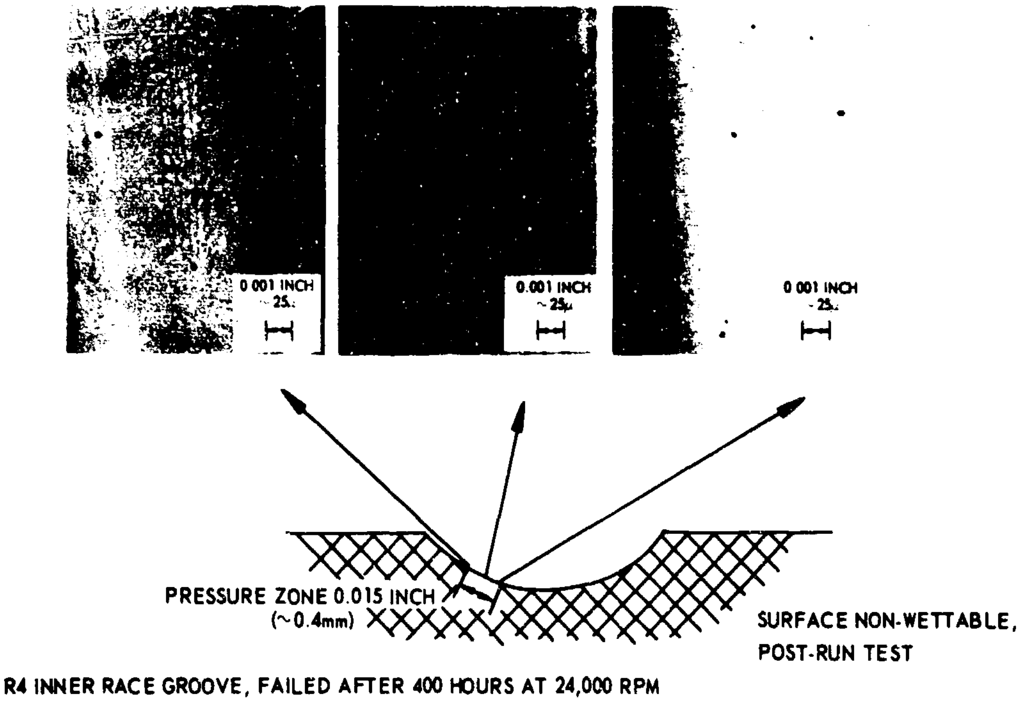
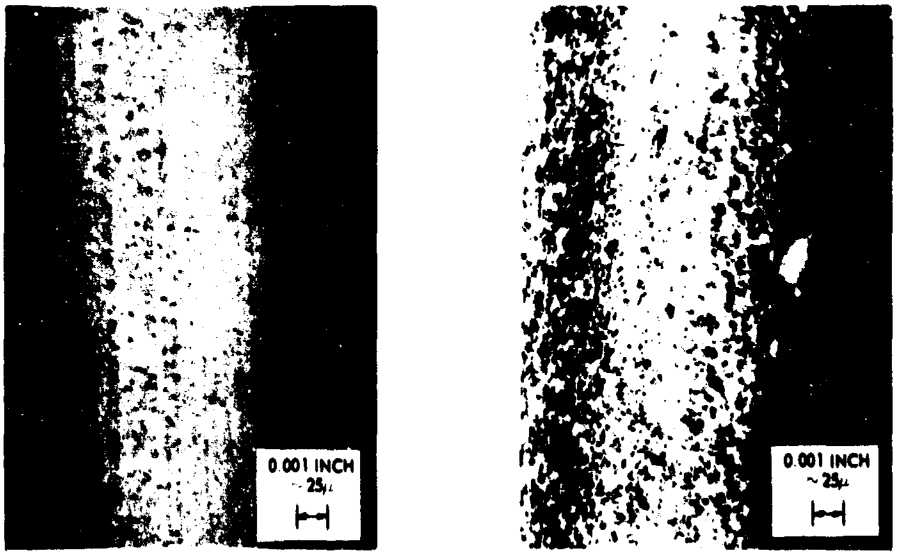


Fig. 16 Chemically contaminated bearing surface, early failure



1500 RUNNING HOURS

170 RUNNING HOURS

INNER-RACE GROOVES
RUN AT 24,000 RPM

SURFACE TESTED WETTABLE
AFTER RUNNING

SURFACE TESTED NON-WETTABLE
AFTER RUNNING

Fig. 17 Surface changes, early bearing failure

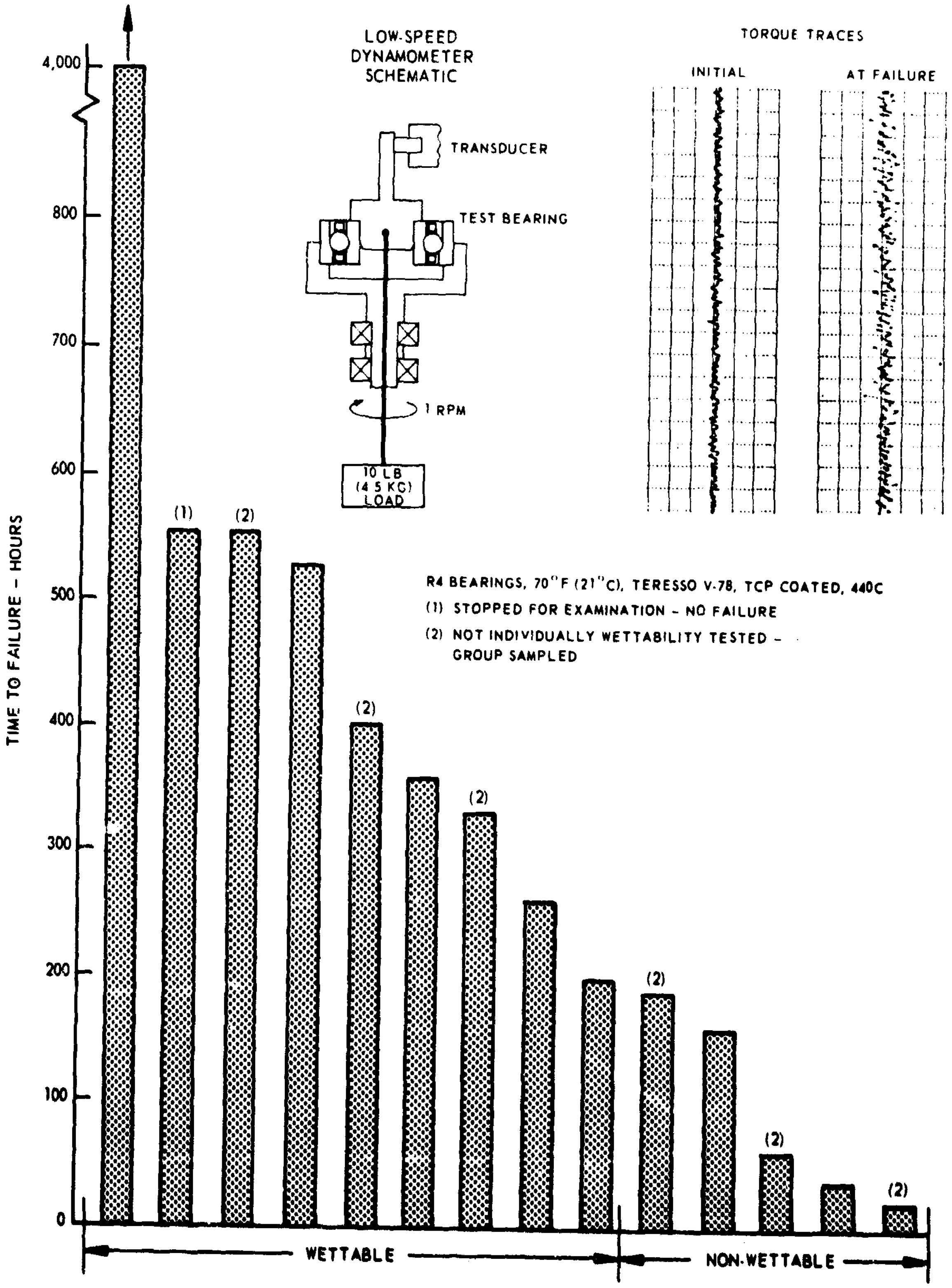
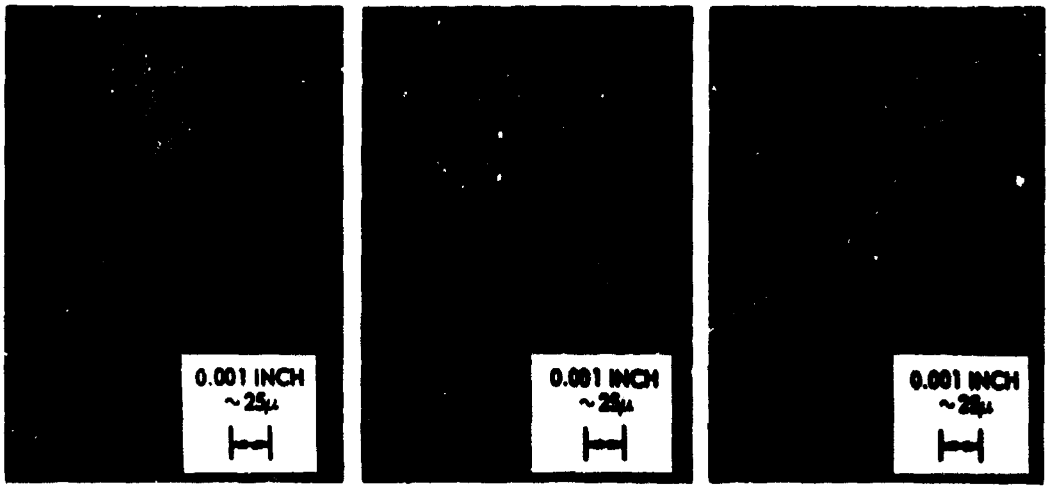


Fig. 18 Surface chemistry effect on low-speed endurance (1-rpm continuous)



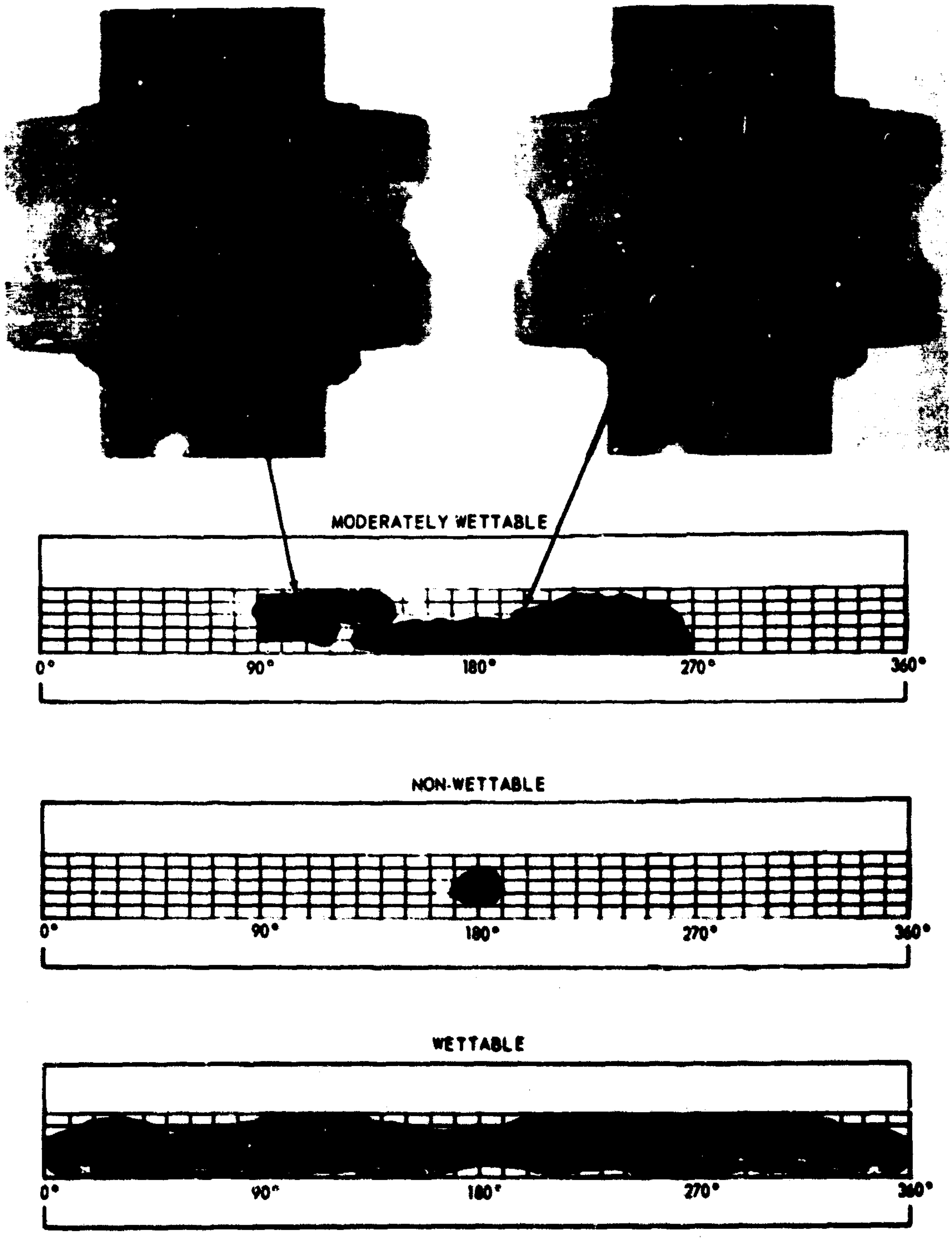
AS RECEIVED

TCP PRE-RUN
SURFACE NON-WETTABLE

TCP PRE-RUN
SURFACE WETTABLE

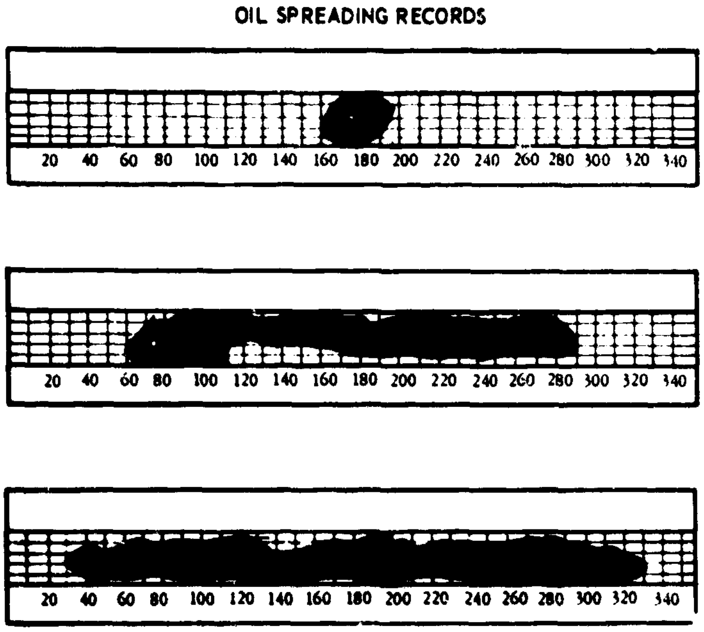
R4 BEARING PRERUNNING CONDITIONS IN FILTERED RECIRCULATING TCP
PRELOAD: 14 LBS (6.4 KG) SPEED: 30 RPM TEMP. 130°F (54°C) TIME OF RUN: 100 HRS

Fig. 19 Surface changes, inner-race grooves, tricresyl phosphate (TCP) prerunning



NOTE: OIL DROP SPREADING TIME 19 HOURS

Fig. 20 Oil spreading records, inner races



AS RECEIVED

AGING TIME AT 180°F (82°C)	EQUIVALENT TIME* AT 70°F (21°C)
1 WEEK	1¼ YEARS
2 WEEKS	2½ YEARS

SURFACES AGED BY EXPOSURE TO TERESSO V-78 OIL AT 180°F (82°C)

*EACH WEEK AT 180°F (82°C) IS EQUIVALENT TO 1¼ YEARS AT 70°F (21°C), ASSUMING REACTION RATE DOUBLING WITH EACH 10°C RISE

Fig. 21 Aging effect on wettability

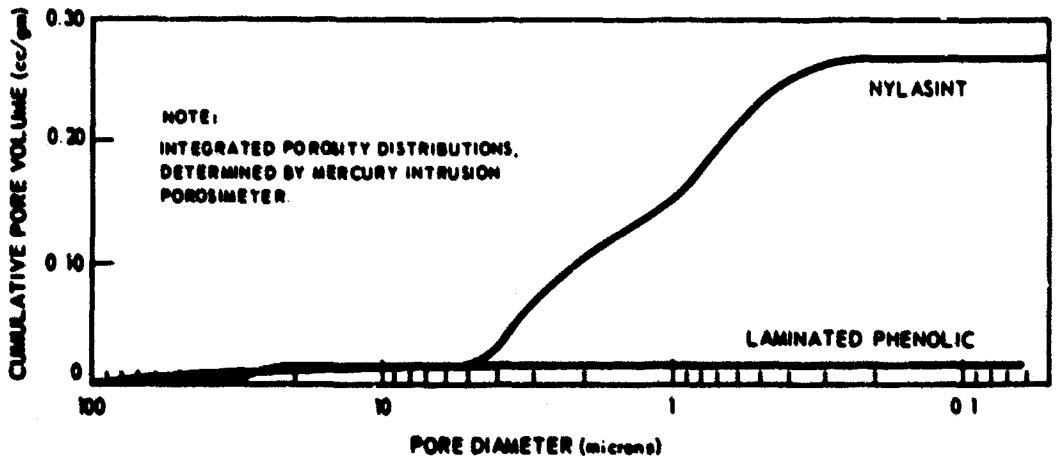
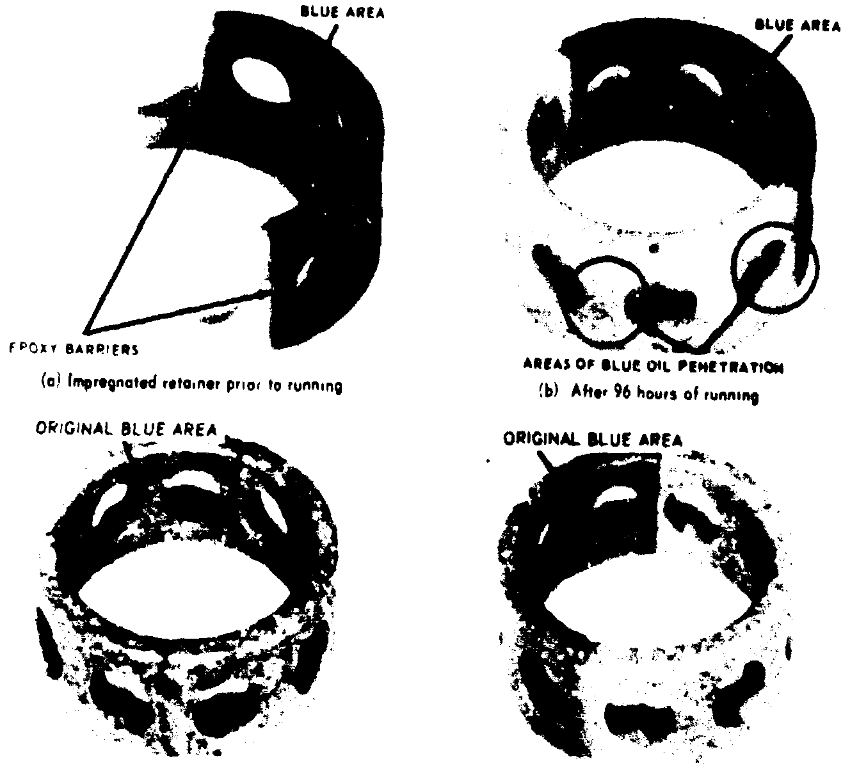


Fig. 22 Pore size distribution for porous nylon and laminated phenolic retainer materials



(a) Impregnated retainer prior to running

(b) After 96 hours of running

NOTE: LEFT HAND RETAINER HAS A LARGER AREA ORIGINALLY IMPREGNATED WITH BLUE OIL, WHICH ACCOUNTS FOR THE DARKER RIGHT HAND SIDE OF THAT RETAINER

(c) After 1786 hours of running

Fig. 23 Oil circulation, Nylasint retainer

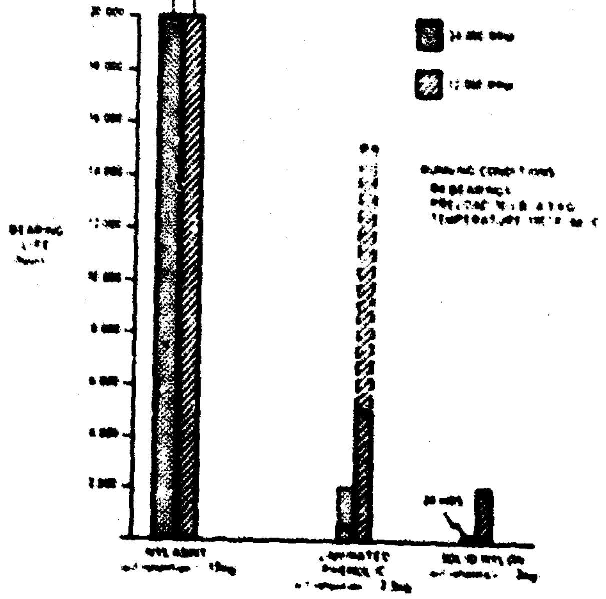
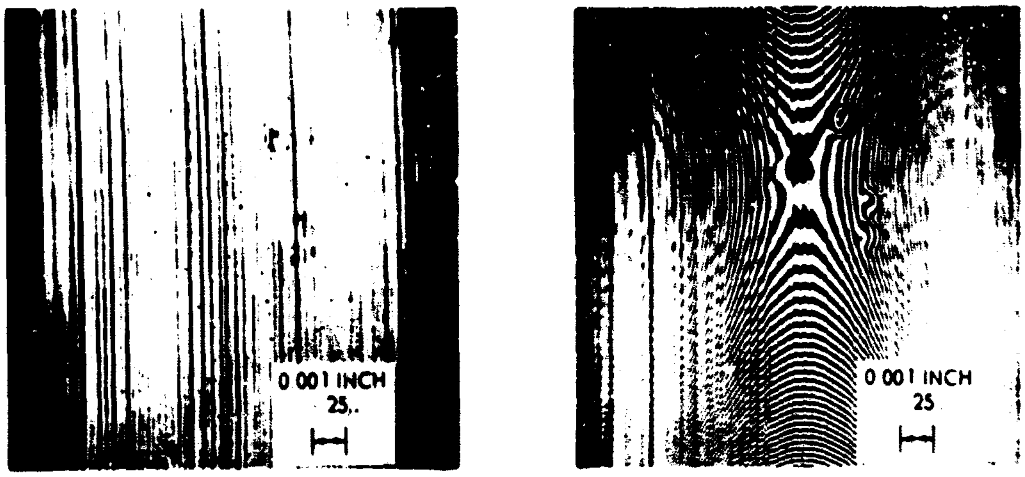


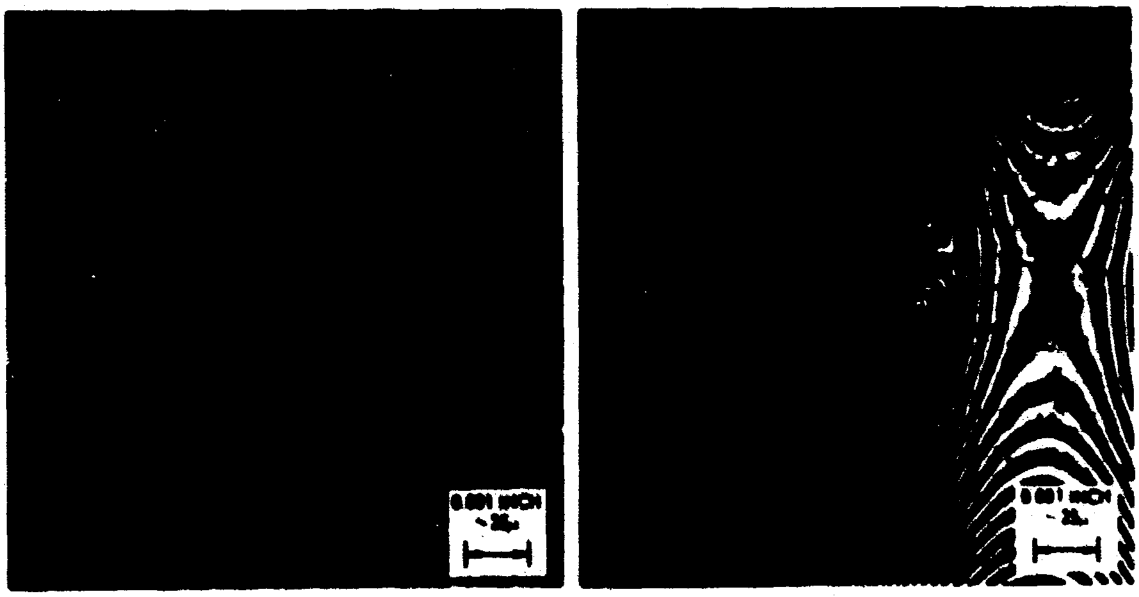
Fig. 24 Expected bearing life versus speed for three retainer materials



NORMAL MICROSCOPIC APPEARANCE

INTERFERENCE FRINGE PATTERN OF SAME FIELD FRINGE SPACING IS 10.6 MICROINCHES (0.27μ) NOTE SHALLOW SIDES OF INDENTATIONS

Fig. 25 Brinells on inner race-groove surface caused by soft particles



NORMAL MICROSCOPIC APPEARANCE

INTERFERENCE FRINGE PATTERN OF SAME FIELD FRINGE SPACING IS 10.6 MICROINCHES (0.27μ) NOTE STEEP SIDES OF INDENTATION

Fig. 26 Brinell on inner race-groove surface caused by hard particle

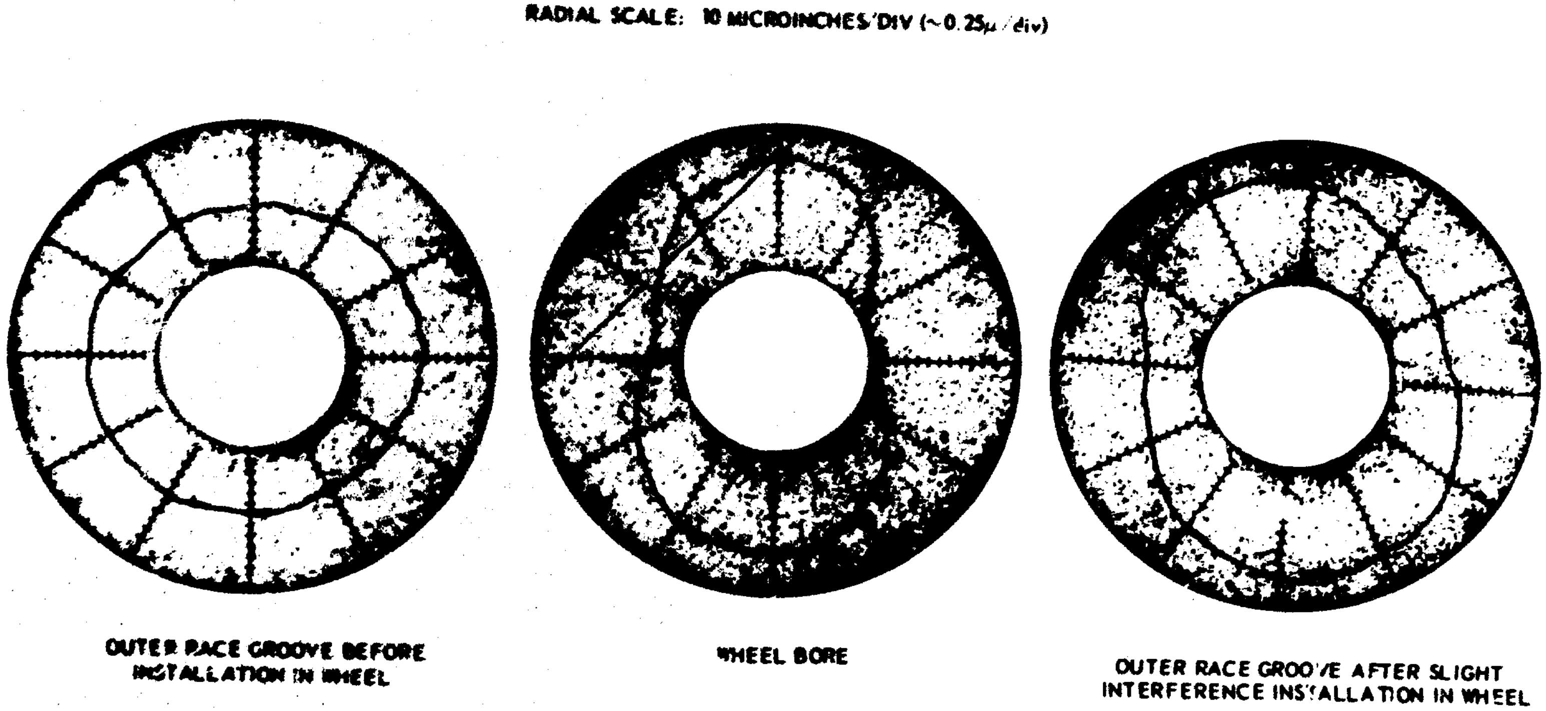
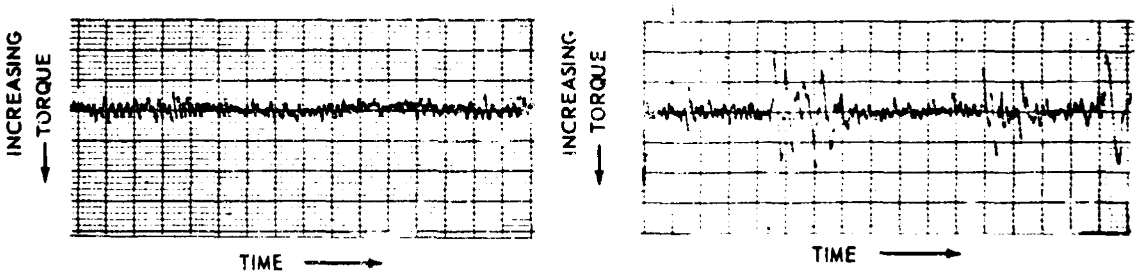
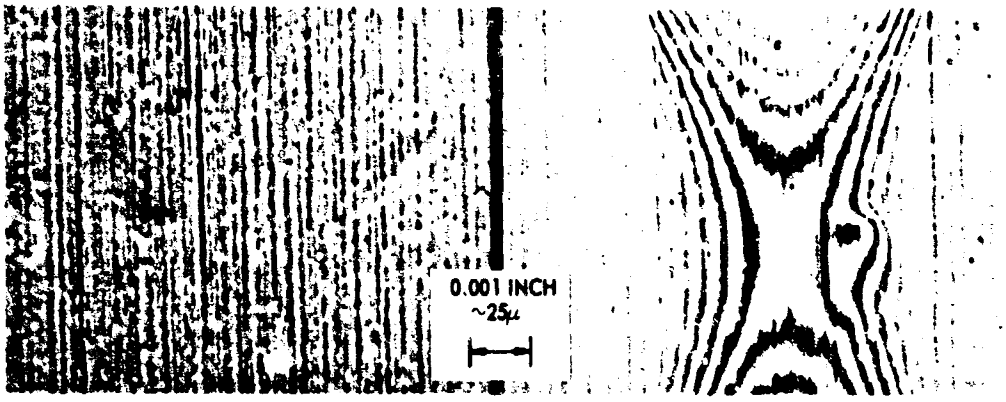


Fig. 27 Distortion of race groove by interference fit to out-of-round member



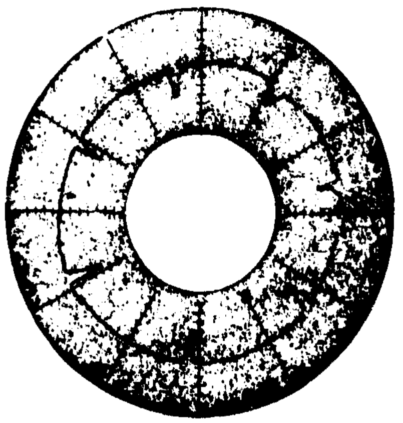
INITIAL LOW-SPEED DYNAMOMETER TRACE - ACCEPTABLE BEARING

SUBSEQUENT LOW-SPEED DYNAMOMETER TRACE - METAL DAMAGE



NORMAL MICROSCOPIC APPEARANCE METAL DAMAGE DIFFICULT TO DETECT

INTERFERENCE FRINGE PATTERN: BRINELLS IN INNER-RACE GROOVE READILY APPARENT. FRINGE SPACING IS 10.6 MICROINCHES (0.27μ).



ROUNDNESS TRACE - 7 BALL INDENTATIONS IN INNER RACE

Fig. 28 Detection of metal damage during bearing processing

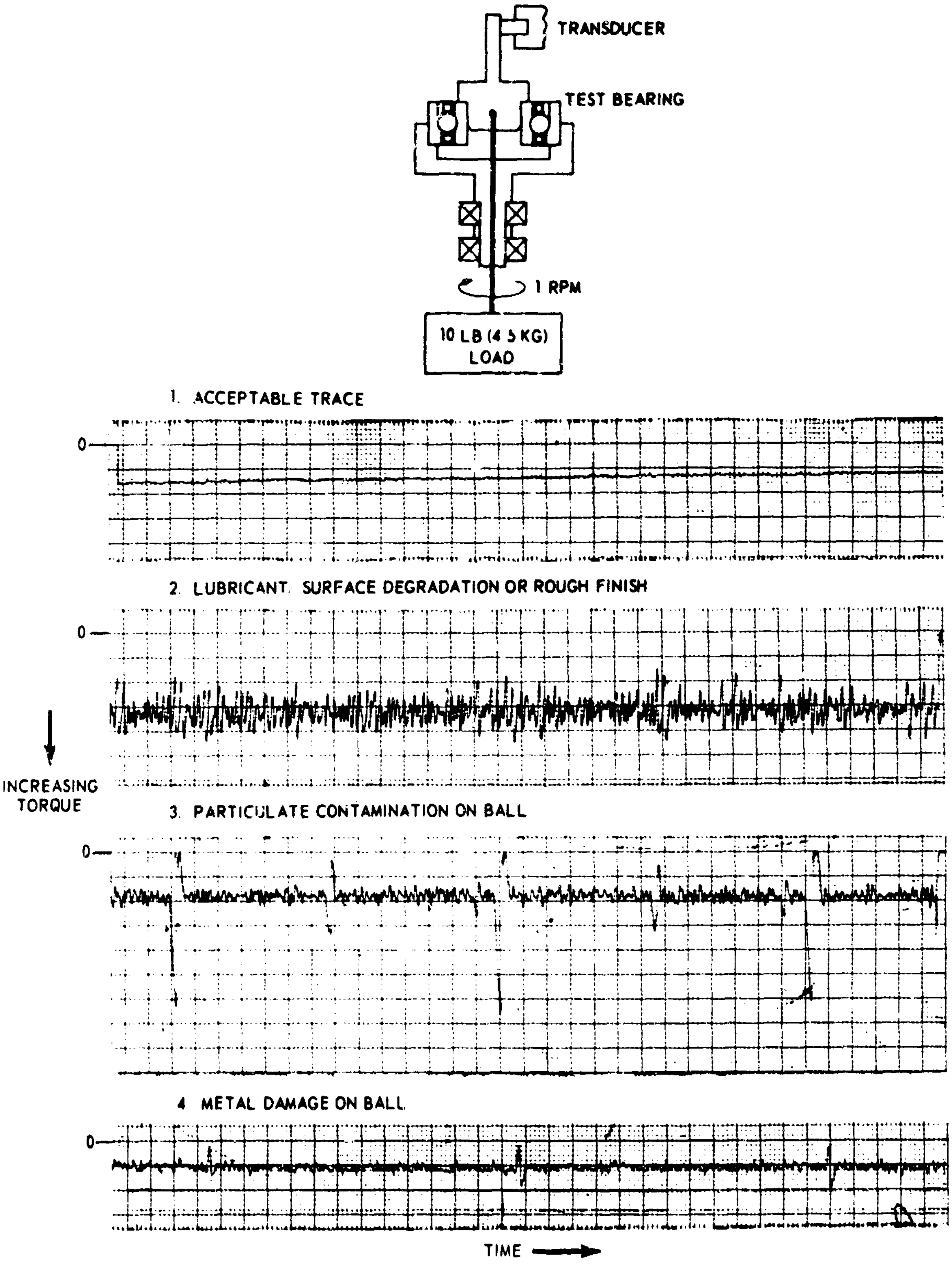


Fig. 29 Low-speed dynamometer and typical traces

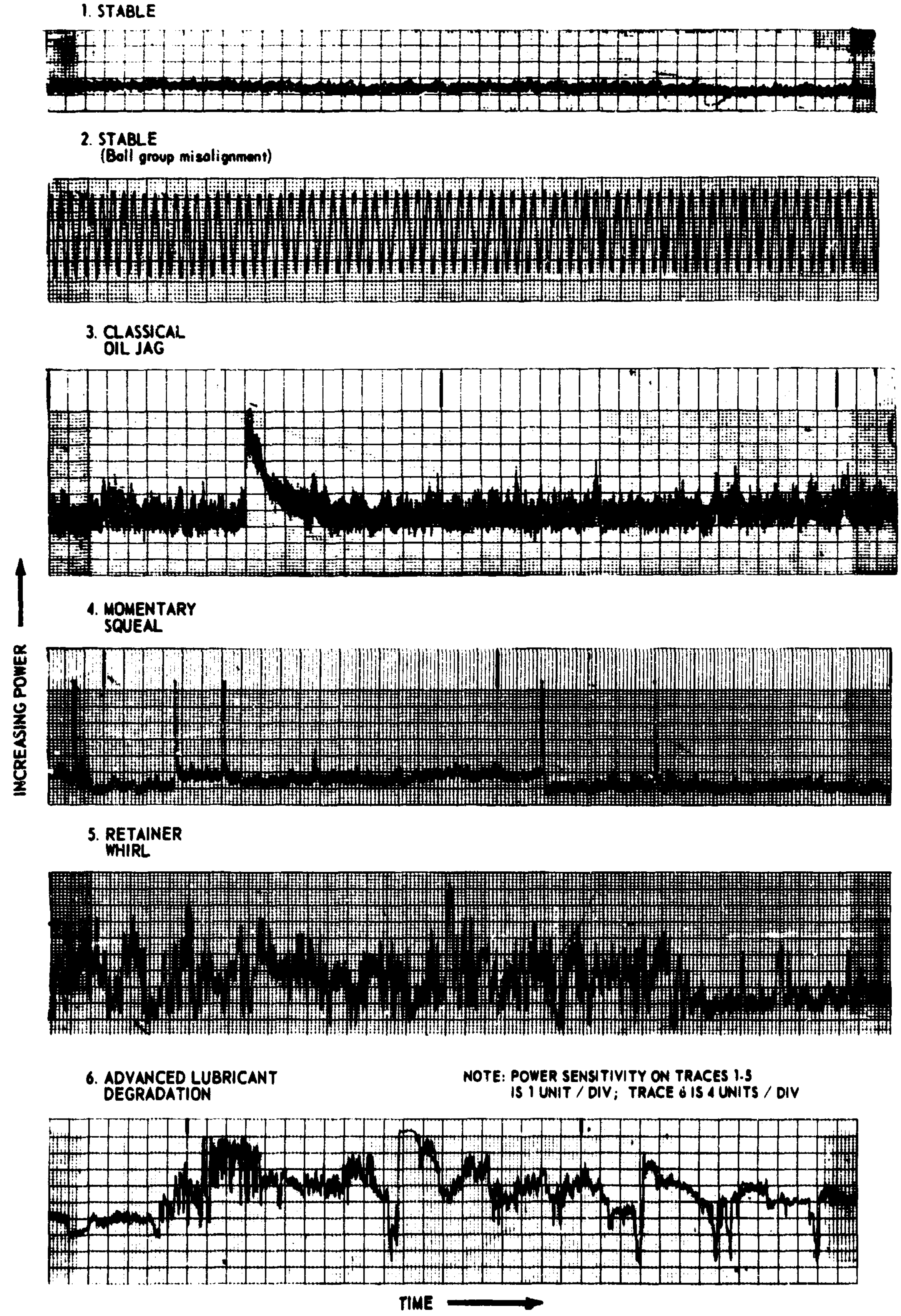
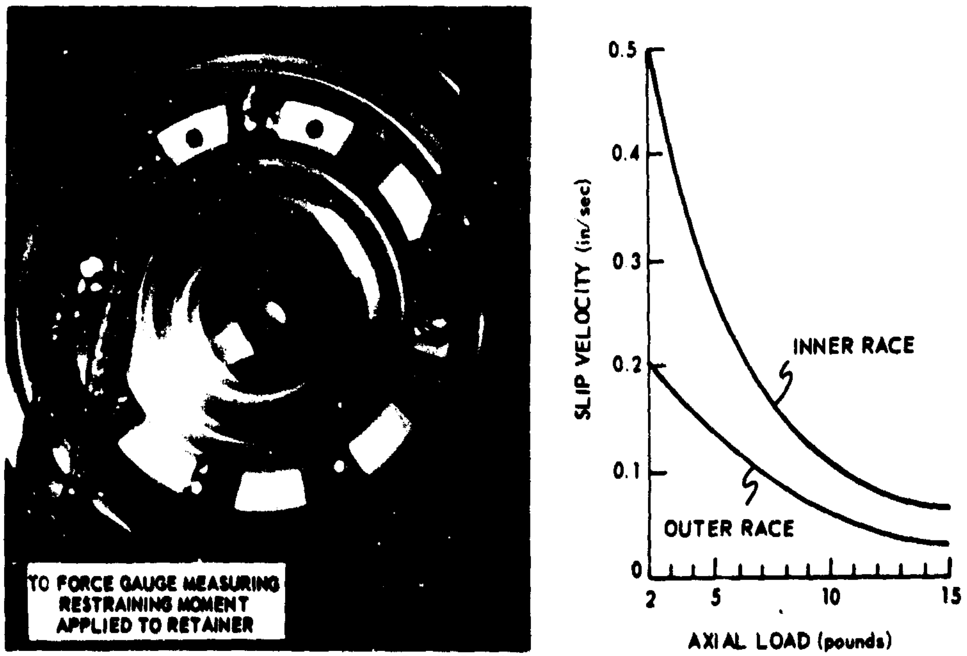


Fig. 30 Bearing behavior as seen on milliwattmeter

EVALUATION TECHNIQUE	SINGLE BEARING	BEARING PAIR TEST	FINAL WHEEL	SEALED FLOAT	GYRO
OPTICAL	●	●			
LOW-SPEED DYNAMOMETER	●	●	●		
MILLIWATTMETER		●	●	●	●
DECELERATION TEST		●	●	●	●
FREQUENCY ANALYSIS		●	●	●	●
WHEEL PACKAGE BALANCE STABILITY TEST			●	●	●
GYRO TEST					●

Fig. 31 Bearing evaluation at successive processing stages



R4 BEARING, RACES COUNTER-ROTATING, BALL GROUP STATIONARY

Fig. 32 Slip of ball to race groove, with fixed retainer moment

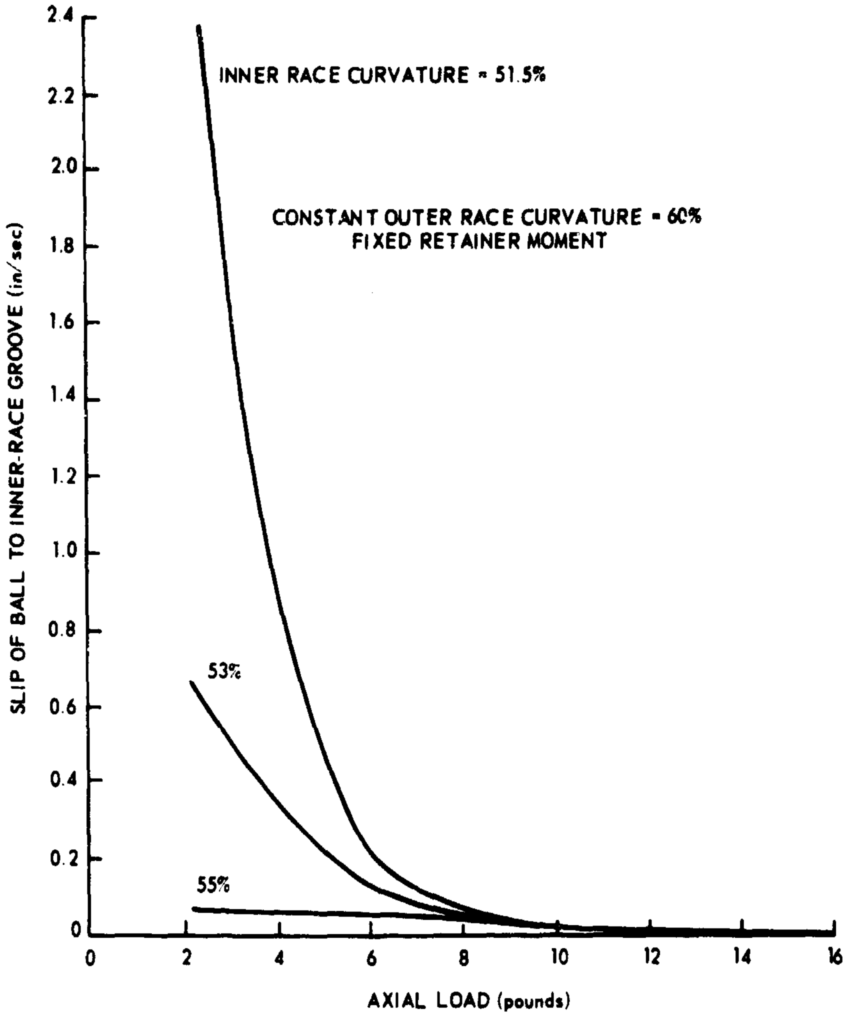


Fig.33 Slip of ball to inner race versus axial load for different inner race curvatures

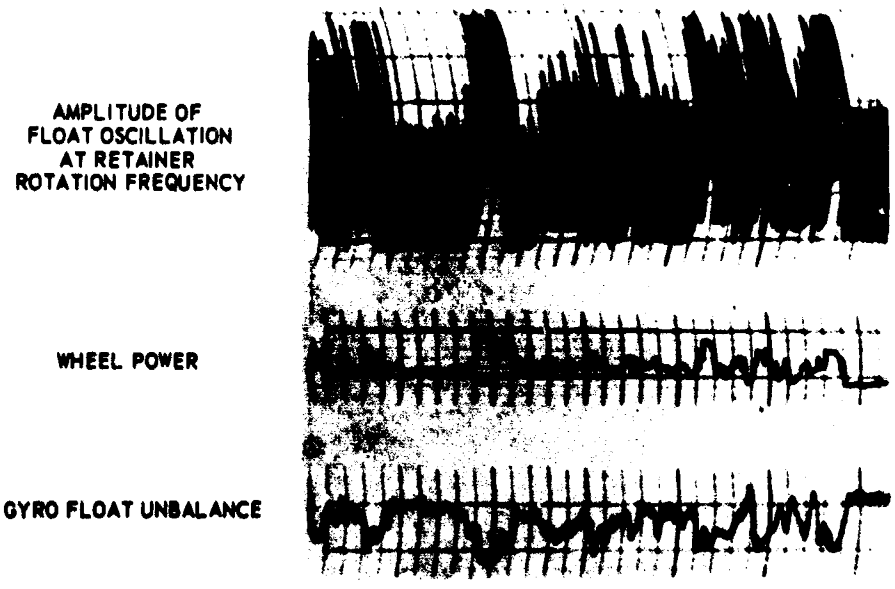


Fig.34 Correlation of gyro unbalance with wheel power and with retainer and ball group motion

CONICAL GAS BEARINGS FOR
GYROSCOPE SPIN AXIS SUPPORT

by

D. Faddy and T.L. Ellis

Royal Aircraft Establishment
Farnborough, Hampshire, England

This paper expresses the opinions of the authors and does not necessarily represent the official views of the Royal Aircraft Establishment.

SUMMARY

The problems of the manufacture of three types of self pressurised gas bearings are compared and the manufacture and inspection of an experimental conical bearing is described. A proposed design for a miniature gyroscope is included.

CONICAL GAS BEARINGS FOR GYROSCOPE SPIN AXIS SUPPORT

D. Faddy and T.L. Ellis

1. INTRODUCTION

At the Royal Aircraft Establishment three types of self-pressurised gas bearings have been considered for spin axis support of inertial quality gyroscope rotors. They are: opposed hemispherical, cylindrical journal and flat thrust plate ("H" form) and conical bearings. The opposed hemispherical bearing and the conical bearing have been investigated entirely at RAE and the "H" form bearing has been the subject of a Government contract with the British Aircraft Corporation.

At first it appeared that considerable difficulties lay in the design of bearings to provide adequate gyroscope performance, particularly in respect of stability and vibration capacity. The work on the stability of bearings both in the USA and in the UK has resulted in satisfactory solutions to this problem in many cases. Further, the actual shape (whether hemispherical, "H" type or conical) is much less important than the projected area of the bearing in the direction of the load and it can be assumed, with certain minor restrictions, that bearings for gyroscope spin axis support with satisfactory static and dynamic performance can be made in any of the three configurations mentioned.

A considerable number of papers have now been published on the theoretical aspects of self-pressurised gas bearings but, apart from the problem of bearing surface materials, relatively little has been published on the practical aspects of manufacture. Considerable difficulties were experienced with the manufacture of hemispheres, which seriously inconvenienced and delayed experimental work. Cylindrical bearings have likewise proved time-consuming and expensive to manufacture. It has seemed to us that the difficulty and cost of manufacturing bearings is delaying their development, not only for gyroscope use, but also for the other applications for which they are theoretically suitable.

We concluded that it would be best to concentrate our efforts on the type of bearing that would minimise the manufacturing problem, both at the development stage and, if possible, in production.

The other problem which has seriously impeded the development of self-pressurised gas bearings is that of the development of suitable materials for satisfactory start-stop and touchdown performance. In view of the effort devoted to this aspect of the subject elsewhere we have, for the time being, ignored this problem and relied on the advice of organisations specialising in this field.

2. COMPARISON OF THE DIFFICULTIES OF MANUFACTURE OF THREE TYPES OF BEARING

2.1 General Requirements

The three types of bearing are shown diagrammatically in Figure 1. The ease with which a bearing can be manufactured cannot be assessed accurately without detailed knowledge of its dimensions and a number of manufacturing trials. However, it is assumed for the purposes of this comparison that the six principal conditions to be satisfied are as follows:

(a) A minimum of highly skilled handwork, such as lapping, should be required.

(b) The number of close manufacturing tolerances should be reduced to a minimum. They are always easier to achieve on a machine than by hand.

(c) Final inspection to gas bearing tolerances is time-consuming and expensive and should be minimised. Intermediate inspection is even more expensive and should be eliminated.

(d) It is difficult to obtain a good surface finish (1-3 microinches) at the same time as a tight dimensional tolerance (10 microinches). The requirement to do both these things at the same time should be avoided as far as possible.

(e) It should be as easy as possible to manufacture grooves or pockets for supporting thrust loads.

(f) The parts should be simple and robust so that they are easy to clean and assemble.

The choice of the bearing type which best satisfies these conditions is bound to be considerably influenced by the grinding machinery already available, unless time and resources are sufficient to permit the development of special purpose machinery.

Although larger machines were available within the Royal Aircraft Establishment the most convenient one for this work was a small Studer type OB machine, 12 years old, but in quite exceptional condition. The machine was improved by the addition of a British Aircraft Corporation externally pressurised air bearing workhead and now produces work round to within 10-12 microinches.

It should be borne in mind, however, that the accuracy and surface finish that can be obtained from cylindrical grinding machines modified from manufacturers' standard has recently shown signs of marked improvement. It is probable that, within the next five years or so, machines of a much higher accuracy will become available at prices comparable with those of present day top quality machines.

The manufacturing requirements and difficulties of the three types of bearings are now discussed.

2.2 "W" Form Bearings

The manufacture of these bearings involves the following:

(a) Either the shaft or the bore of the journal have to be made to size to within 10 microinches. Normal practice seems to be to grind the shaft between centres and to lap the bore by hand, checking straightness from time to time. It is possible to grind the rotor bore, but this has not usually been done as far as we are aware.

(b) The four thrust faces have to be made perpendicular to the axes of the bore and shaft to a very small tolerance. Three arc seconds is typical on miniature bearings.

(c) The length of the shaft and bore relative to each other have to be controlled to an accuracy of 10 microinches in order to control thrust plate clearance.

(d) In order to suppress half speed whirl, either the bore has to be accurately mutilated by lobing or the shaft has to be helically grooved.

(e) The thrust plates have to be spiral grooved. This is a highly developed operation which is probably easier to do on flat thrust plates than on any other type of surface.

Of this list all stages, except possibly (e), require hand operations and/or intermediate inspection.

2.3 Opposed Hemispheres

This arrangement has the unique advantage for single axis support that no alignment of the bearing surfaces is necessary. It has also, in common with conical bearings, the advantage of stability as the grooves necessary for thrust support also serve to suppress whirl. The manufacturing difficulties are as follows:

(a) It is necessary to make the males and females accurately spherical to within 10 microinches and one of them to size within 10 microinches.

(b) A straightforward method of measuring the diameters of female hemispheres other than by glass masters is required.

(c) Spiral grooves are more difficult to make in hemispheres than in any other type of surfaces. Pockets are much easier but give considerably reduced performance compared to grooves.

Considerable difficulty has been experienced with the manufacture of hemispheres for experimental purposes at RAE. They have been made mainly by hand lapping to glass masters. This is a time-consuming and expensive operation and it is difficult to get good operators to do it for protracted periods. An automatic lapping machine has been made but, although satisfactory for portions of spheres away from the polar and equatorial regions, it has not given good results for complete hemispheres.

Plunge grinding appears to be a much more promising method of producing female hemispheres economically, but really requires a special purpose cylindrical grinding machine with precise control of wheel or workhead centre height.

The Instrumentation Laboratory of the Massachusetts Institute of Technology have published the results of a considerable amount of work on bearings of this type[1] and have developed equipment for measuring female hemispheres and for machining spiral grooves. The methods used by them appear to be expensive and time consuming.

2.4 Conical Bearings

The basic advantage that conical bearings possess is that none of the bearing surfaces have to be made to size. The bearing film thickness can be controlled by adjusting the length of a separately manufactured spacer or something similar. This means that an acceptable surface finish is much easier to obtain from a grinding machine. They are also almost as easy to groove as are flat thrust plates.

Conical bearings share with hemispherical bearings the advantage already mentioned that grooves for thrust support will also serve to give stability to the journal and also that female surfaces can relatively easily be coated with hard materials.

The tight tolerances required are as follows:

(a) It is necessary to hold the cone semi-angle of mating male and female cones to within 8-10 arc seconds of each other on bearings of about 0.4 inch generator length. This appears to be relatively easy to do on a suitable grinding machine by suitably ordering the machining operations.

(b) The axes of the four bearing surfaces have to be held coincident with each other to a high accuracy. This is much the most difficult part of the manufacture of conical bearings but, by careful use of mandrels (see Section 3), bearings can be aligned to sufficient accuracy without intermediate inspection. Further, this problem can be considerably reduced in some bearing designs.

(c) One male cone must be a very precise sliding fit onto a central shaft in order to maintain alignment while allowing adjustment of the clearance.

2.5 Comparison

The main points of comparison detailed above are summarised in Table I, which indicates fairly clearly that the conical design has many advantages. It was on this sort of reasoning that we decided to concentrate effort onto the conical bearing.

At the time that this decision was made it appeared that conical bearings offered a simplicity of manufacture which had immediate advantages for experimental work and longer term potential for cheaper production bearings. It was appreciated that other organisations had experimented with this type of bearing and finally used some other form in production but, in the absence of published information as to the reason for this, it was felt that this information had to be ignored.

3. MACHINING METHOD FOR TEST BEARING SURFACES

The problem of manufacturing conical bearing test pieces is basically that of ordering the machining operations so that the requirements of matched cone angles and alignment are achieved without any sophisticated setting-up procedure.

The materials chosen for the test pieces were Stellite and an air hardening tool steel, for the male and female cones respectively. The latter material was chosen because spiral grooves can be easily etched into the surface.

The bearing parts (Fig.2) were machined to normal machine shop tolerances, and the female portion was heat treated. The holes in the male cones were accurately ground and lapped to become a firm fit onto a plug gauge. The mandrel in the workhead (Fig.3(a)) was also ground to provide the same type of fit into the male cone bore. The workhead was set at the required angle and the male cones were ground to the required finish (Fig.3(b)).

It is most important that the mandrel should not be moved until the male cone grinding is completed, and that the workhead should not be moved until all the conical surfaces are finished. The cylindrical mandrel was replaced with a conical type (Fig.3(c)) and ground to the same angle; the rotor was clamped into position (Fig.3(d)) and the exposed cone was machined. The rotor was then reversed on the mandrel and the other female machined. This procedure should ensure that the axis of the female cones are coincident to a high order of accuracy. Finally the workhead was re-aligned parallel to the traverse, and the outside of the rotor was ground concentric with the female cones.

4. INSPECTION METHODS

The inspection of conical surfaces presents some special problems. The measurements that have to be made are as follows:

(a) roundness,

(b) surface finish,

(c) straightness of generating lines,

(d) cone angle,

(e) axial alignment of the male surfaces and of the female surfaces.

The first three are standard inspection procedures common to cylindrical bearings and do not need describing. Methods of measuring the other two were developed in association with RAE Inspection Department and are described now.

Although means of measuring conical plug gauges are well known and more sophisticated methods of measuring master gauges have been described recently², none of those known to us were suitable for the inspection of female surfaces. The principal equipments required are an optical dividing head and a Taylor Hobson Mitronic Comparator. The "Mitronic" was originally developed by Ferranti Limited from a gyroscope pick-off and is basically an electronic dial gauge capable of detecting movements of 2 microinches (0.05 micron). The dividing head is placed on a surface table, with its rotating plate set perpendicular to the table surface. The two male cones assembled on their shaft between centres are mounted on the rotating plate with their axis parallel to it; the plate is rotated until a cone generating line is parallel to the surface table, using the "Mitronic" as a height comparator at each end of the bearing surface. The angle of the dividing head is recorded and the procedure repeated for the other three generating lines in the same vertical plane. From this information both cone angles can be deduced, together with the angular misalignment of the cone axes. This procedure must be repeated with the cone assembly rotated through a right angle to establish the spatial relationship of the cone axes. The angular alignment of female conical surfaces can be measured similarly.

There remains the problem of establishing the perpendicular distance between the cone axes at one point. This can be done for the rotor, by rotating it with one female surface separated from a mating cone by a thin film of oil and measuring the total indicated run-out of the unsupported female cone with the "Mitronic" comparator. The cones can be rotated with a simple motor, through a friction drive, and results repeatable within 5 microinches (0.08 micron) can be obtained. A similar system can be used to measure the male cone assembly.

5. GROOVE MANUFACTURE

The methods that have been investigated at RAE are: electro-chemical etching, chemical etching and grit blasting. Of these, electro-chemical etching has now been abandoned because it is considered unnecessarily complicated for gas bearing purposes. Chemical etching is the simplest of the two other methods, but cannot be used for ceramic surfaces. Both methods are suitable for development as inexpensive production methods.

A flat master plate is etched with grooves and used to project the pattern onto the bearing surface which has previously been coated with a positive light sensitive varnish (Fig.4). Ultra violet light is used and, after development, the unexposed areas of varnish are washed away. The part is then submerged in dilute acid for a predetermined time to etch the grooves to the required depth. Several combinations of nitric, hydrochloric and sulphuric acids have been tried on hard steel. The most consistent results have, however, been obtained with a 1% solution of nitric acid in distilled water, heated to 60°C and vigorously agitated. Oxide residue in the bottoms of grooves can conveniently be removed with a hard pencil eraser leaving a clean finish with sharp edges, but with a slight tendency to under cut.

For ceramic bearings a grit blasting technique has been developed. For these, conical shell stencils (Fig.5) are made by the same process as described above, except that the groove pattern is etched right through the stencil. A small grit blasting machine manufactured by S.S.White and using fine aluminium oxide powder has been adapted to make the grooves (Fig.6). The male cone, covered by the closely fitting mask is rotated about its axis at about 200 rev/min. The nozzle of the grit blasting machine is mounted on an arm so that it is swept back and forth along the surface of the cone as it rotates. In this way, the grit is distributed uniformly over the area to be grooved. The main difficulty that has so far been experienced is that, in order to achieve a tolerance of about ±20 microinches on groove depth, extreme uniformity is required of the bearing material. This is difficult to obtain with flame plated coatings, unless special precautions are taken.

6. RESULTS OF SOME MANUFACTURING TESTS

Three rotor and bearing assemblies of the type shown in Figure 2 were manufactured by the method described in Section 3. The first of these was made on an unmodified Studer OB machine and the others were made on the same machine after it had been modified by the addition of the externally pressurised air bearing workhead. They were inspected by methods which vary in detail from that proposed for future bearings and outlined in Section 4.

The results are given in Table II and Figure 7. The total indicated run-out measurements are of the outside edge of one female surface relative to the other and of the shaft where it protrudes from the male cone relative to the male surface. These figures include the error in roundness. The straightness tolerance is measured over 90% of the generator length, so that edge effects as the grinding wheel begins and ends its cut are ignored.

These results are too few to be entirely convincing, but indicate that this type of bearing can be satisfactorily made on a suitable high precision grinding machine. The most difficult problem is alignment of the conical surfaces and in this respect the chosen manufacturing test piece was a severe test. It is also possible to positively lock the grinding machine workhead at the required angle with a removable taper peg. A simple experiment done recently suggests that this could enable the rigid order of machining operations imposed by the present method of manufacture to be abandoned. A further possibility, arising from this, is the manufacture of shell bearings which could be attached to less accurately made conical backing pieces with a filled resin during assembly. Alignment of the shells would then be enforced by the mating parts.

Although the surface finish achieved on hard steel and Stellite has been satisfactory, a considerable amount of time has been required to achieve it. Recent grinding tests on flame plated ceramics with diamond wheels have resulted in rather unsatisfactory finishes. This, it is felt, is a result of trying to work too close to the limits of the grinding machine.

7. EXPERIMENTAL RESULTS

A flywheel was added to the second rotor assembly and its radial and axial compliances were measured at a range of speeds up to 300 cycles/second (Fig. 8). These are plotted in Figure 9, where details of the bearing are also given.

The radial compliance should be the same as the axial compliance in a satisfactory gyro spin axis bearing, and in this respect the results are not good. However, they are of interest because the attitude angle of the bearing is shown to be very small at speeds over 200 cycles/second. This would result in a very low cross anisoelastic drift rate, even though the radial compliance is very high.

The groove angle and groove depth of this bearing are much too high and experimental bearings of more realistic design are being manufactured at present.

8. A PROPOSED DESIGN FOR A SINGLE AXIS GYROSCOPE

We have considered as a design exercise, a suitable gas bearing for an inertial quality miniature gyroscope to meet the following requirements:

(a) Total weight of wheel assembly, including motor but excluding float: 50 gm.

(b) Angular momentum: 10^5 c.g.s. units.

(c) To fit a cylindrical float of internal diameter: 3.6 cm.

Two basic possibilities of external and internal motor stators were considered. The external stator has the advantage of better heat dissipation and greater simplicity of bearing design. The internal stator, on the other hand, offers the advantage of largest angular momentum for a given rotor weight. In fact the decision to use an internal stator arrangement was taken on the practical grounds that a suitable motor had already been developed by the British Aircraft Corporation for an "H" form design and we know of no suitable existing external stator design for our purpose.

Two possible alternative materials were considered. The first was to make the assembly (excluding motor parts) from solid ceramic. The ceramic of which we have most experience is Sintox manufactured by Smith's Industries Limited. Although it is of high porosity, which results in a reduced hydrodynamic performance, it has been shown by the British Aircraft Corporation to have a satisfactory start-stop life. Ceramics have, unfortunately, a very low thermal conductivity, which makes them suspect for internal stator designs. The design illustrated in Figure 10 therefore makes use of the second possibility, i.e. flame plated aluminium alloy, as the rotor and bearing material. The thermal properties of this material are good.

The bearing material to be deposited by flame plating is still the subject of experiment but will probably be aluminium oxide, although tungsten carbide and chrome oxide are being considered as alternatives.

The main points of the design are as follows:

(a) The rotor female surfaces are aligned by grinding both bearings without moving the rotor in the workhead. In order to do this the grinding wheel has to be carried on a long mandrel to enable the surface closest to the workhead bearings to be machined.

(b) The male cones are of aluminium shrunk onto a central non-magnetic stainless steel tube. One male cone is to be permanently assembled onto the chrome plated stainless steel shaft before finish machining. Clearance is adjusted by lapping the end of the stainless steel tube on which the other cone is mounted. Non-magnetic stainless steel was chosen because its high thermal coefficient of expansion provides a reasonable match with that of aluminium.

9. CONCLUSION

The results achieved so far show the conical type of bearing to be quite promising for gyroscope spin axis use. They have the considerable advantage that they appear to be relatively easy to make on the highest quality of cylindrical grinding machines available. There are a number of possible ways of overcoming the alignment problem which is their most serious drawback.

It should be remembered that we have primarily considered the problem from the point of view of our immediate manufacturing resources. It has become clear to us that it is possible to develop high quality special purpose grinding machines suitable for the manufacture of any type of self-pressurised gas bearing. The availability of such a machine might modify some of the data used in the comparison of the various bearing designs in this paper.

ACKNOWLEDGEMENT

This paper is Crown Copyright and is reproduced by permission of the Controller, Her Majesty's Stationery Office.

REFERENCES

1. McEwen, J.K.	*Gas-Bearing Fabrication and Measurement Techniques*. Proceedings of the Gyro Spin Axis Hydrodynamic Bearing Symposium. Massachusetts Institute of Technology, Instrumentation Laboratory, 1966.
2. Farago, T.	*Angle Measurements of Cone Shaped Technical Bodies*. Microtecnic, Vol.XXI, No.34, 1967.

TABLE I

Comparison of Manufacturing Problems for Three Types of Bearing

Type	*Angular tolerances and alignment of axes*	*Dimensional tolerances*	*Spiral groove manufacture for thrust support*	*Mutilation for whirl suppression*	*Suitability for coating surfaces with hard materials*
"H" type	4 thrust surfaces to be square to the journal axes to 3 arc seconds Very difficult	Diameter of journal or bearing (10 microinches) Length of journal or bearing (10 microinches)	Straightforward and relatively cheap methods available	Lobing usual but requires much hand-work and intermediate inspection Spiral grooves more effective but also difficult to make	Very difficult to coat bores
Opposed Hemispheres	None	Diameter of either male or female hemispheres Difficult metrology or manufacturing problem	Very difficult and probably expensive Pockets much cheaper but not so effective	Not required	Relatively easy to coat both surfaces
Conical	Cone angles to be the same Relatively easy Axes of the 4 bearing surfaces to be accurately aligned Difficult	None (except of a separately manufactured spacer)	Almost as easy as with flat thrust plates (for cones of semi-angle 30° or more)	Not required	Very easy to coat both surfaces (for cones of semi-angle 30° or more)

TABLE II

Results of Some Manufacturing Tests

	Cone semi-angle	*Roundness (microinches MZC)*	*Straightness tolerance of generators (microinches)*	*Surface finish (microinches CLA)*	*Total indicated run-out (microinches)*
Rotor Assembly No.1					
Rotor end A	30° 5' 44"	20	30	2	**Satisfactory method of measurement not developed at this stage.**
Rotor end B	30° 5' 40"	18	30	2	
Male No.1	30° 5' 46"	25	10	2	
Male No.2	30° 5' 55"	25	10	2	
Rotor Assembly No.2					
Rotor end A	30° 1' 44"	10	14	2-3	25
Rotor end B	30° 1' 48"	10	15	2-3	
Male No.1	30° 1' 56"	10	5	2	25
Male No.2	30° 1' 48"	10	10	2	100
Rotor Assembly No.3					
Rotor end A	-	15	14	2-3	30
Rotor end B	30° 7' 40"	15	14	2-3	
Male No.1	30° 7' 58"	10	10	2	15
Male No.2	30° 7' 54"	10	8	2	10

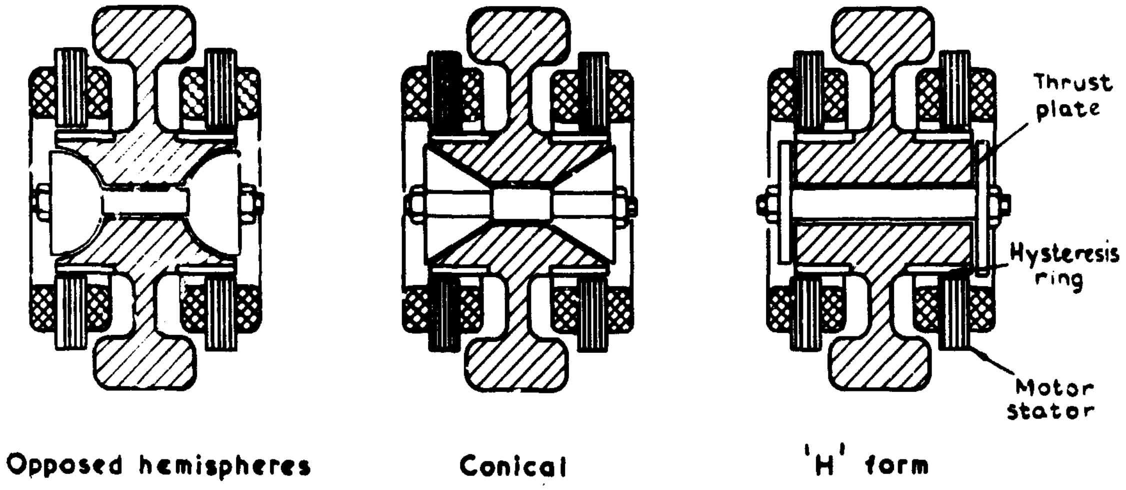


Fig. 1 Types of bearing for the support of gyroscope rotors

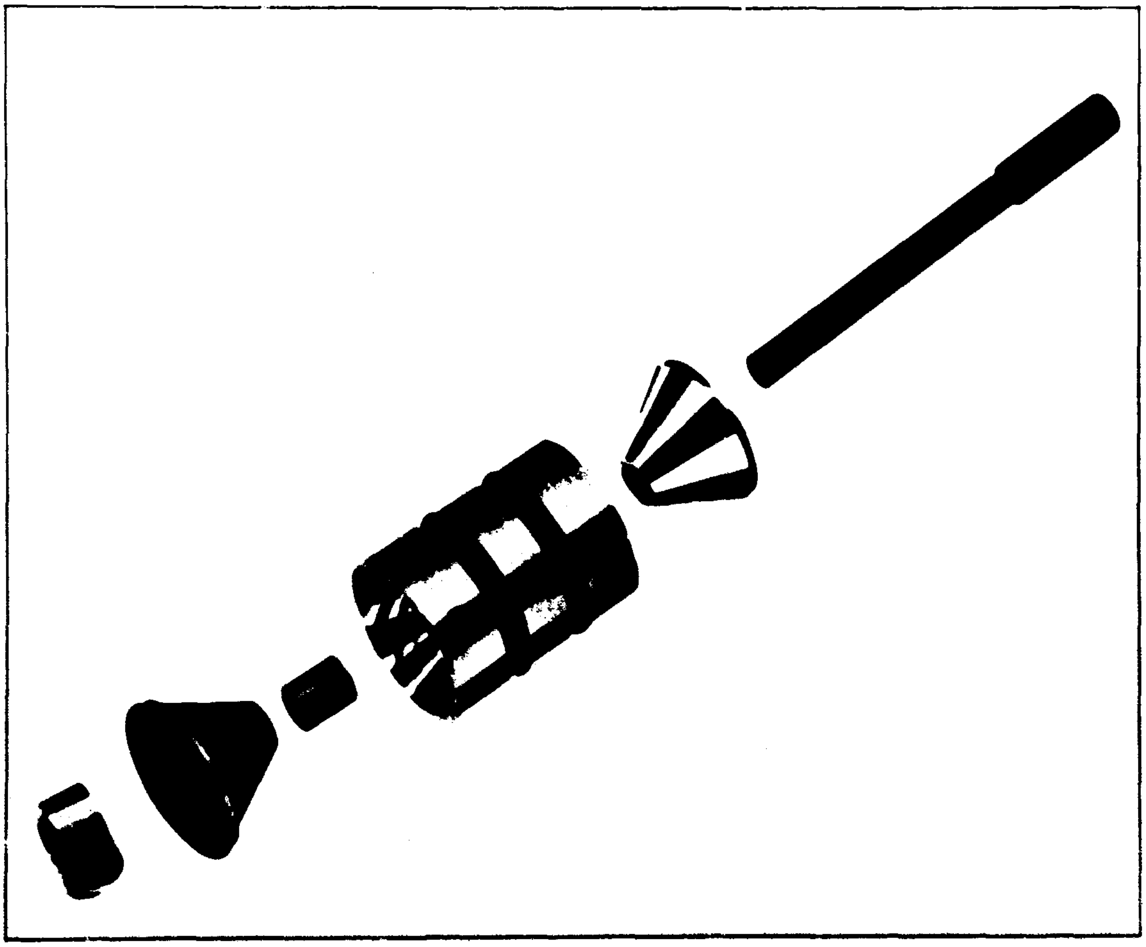


Fig. 2 Rotor and bearing assembly for manufacturing tests

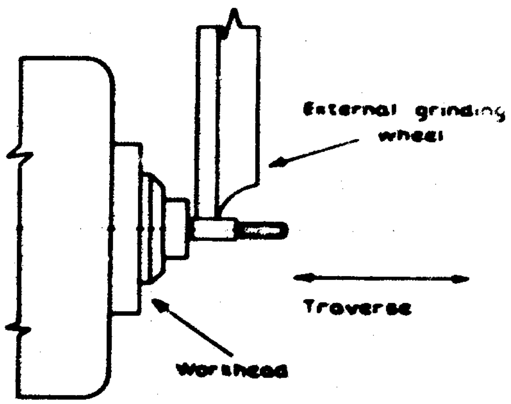


Fig. 3(a) Machining cylindrical mandrel

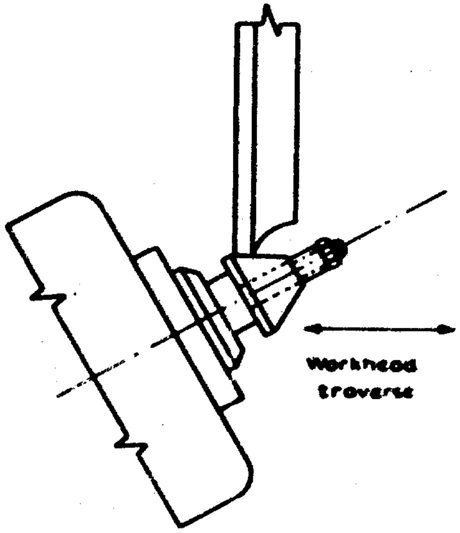


Fig. 3(b) Machining male cones

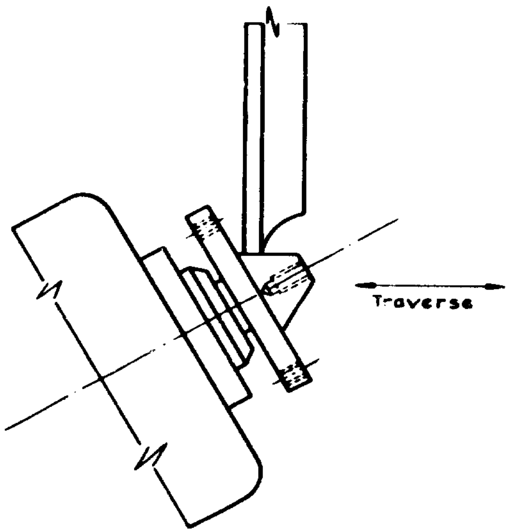


Fig. 3(c) Machining conical mandrel

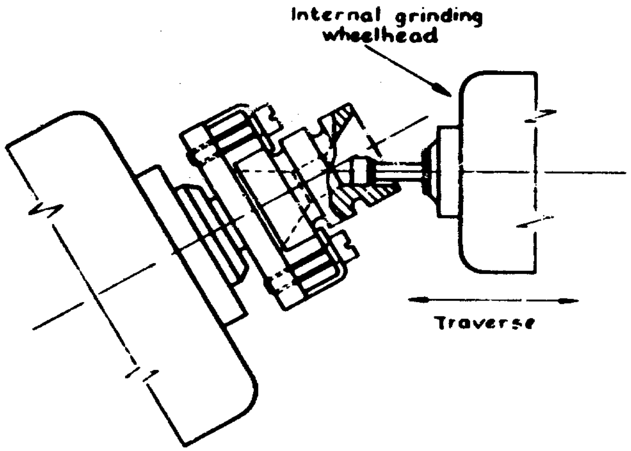


Fig. 3(d) Machining female cones

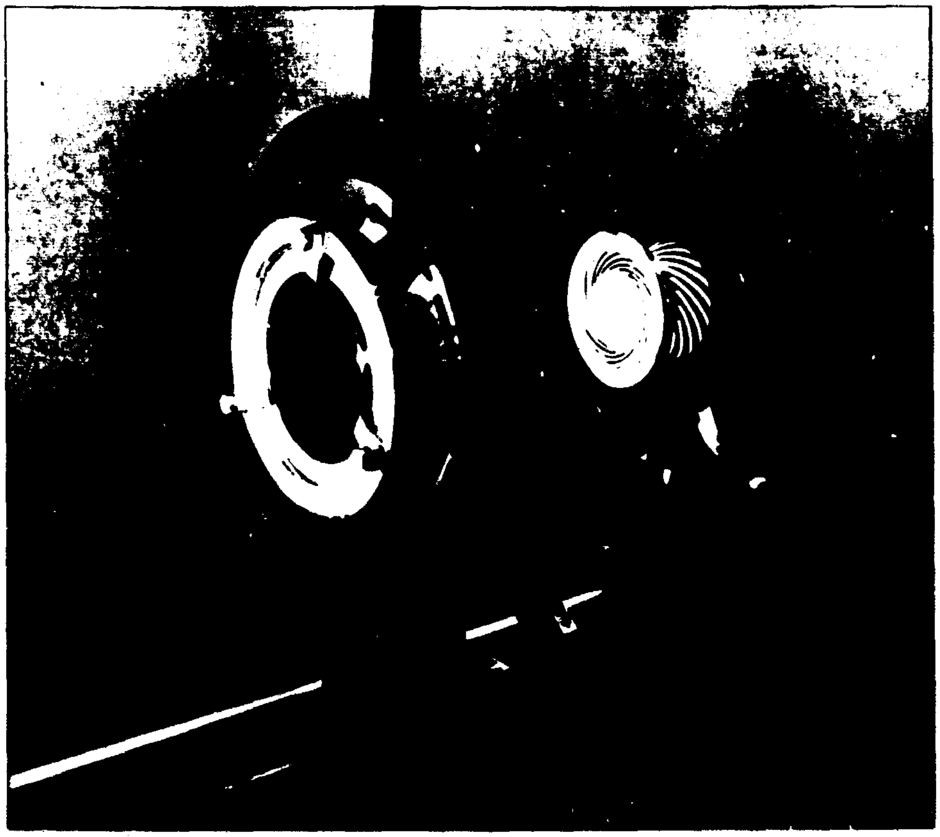


Fig. 4 Projection of groove patterns onto conical shapes

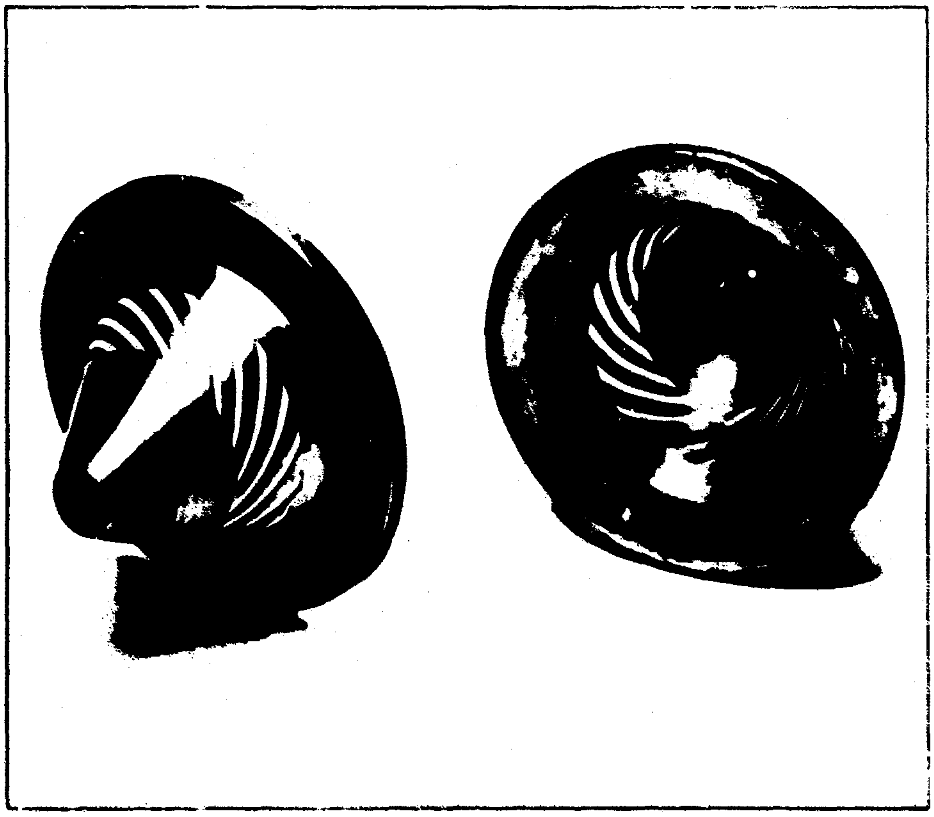


Fig. 5 Shell stencils for grit blasting

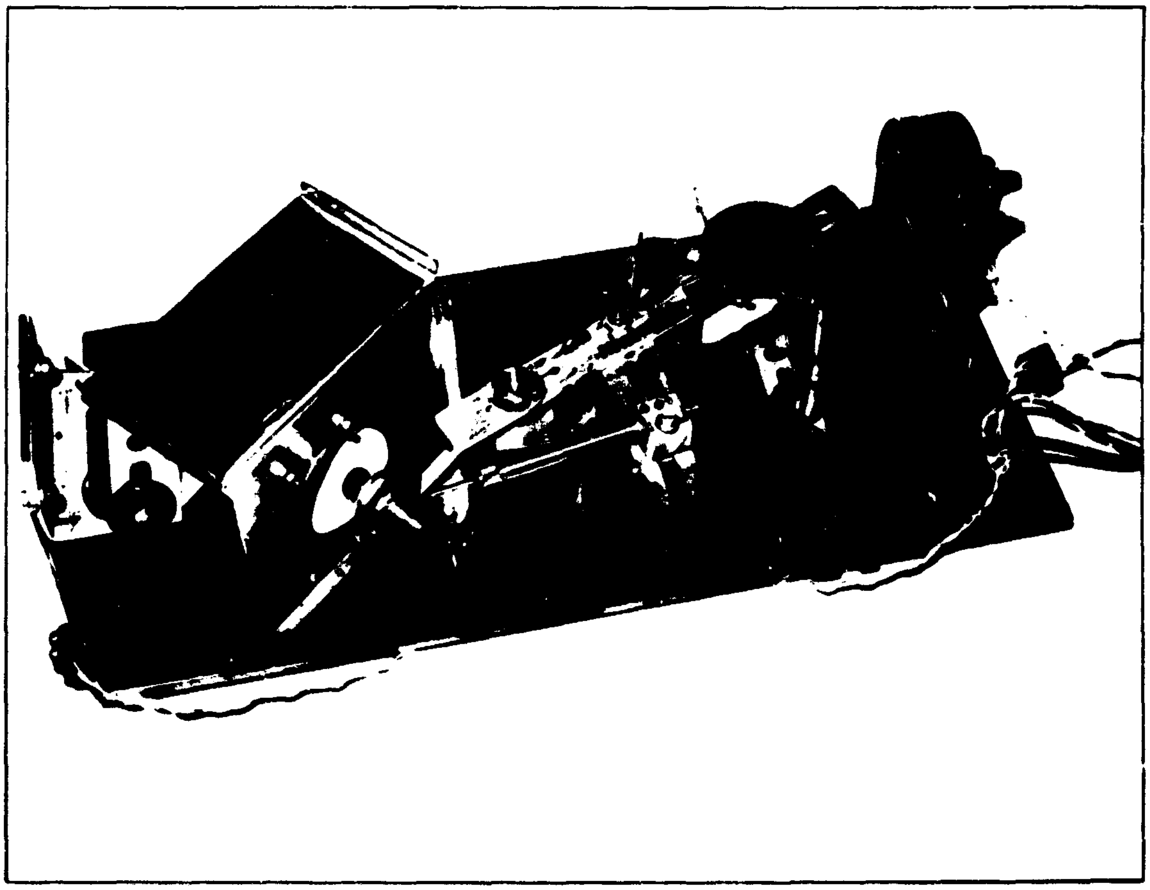


Fig.6 Grit blasting apparatus

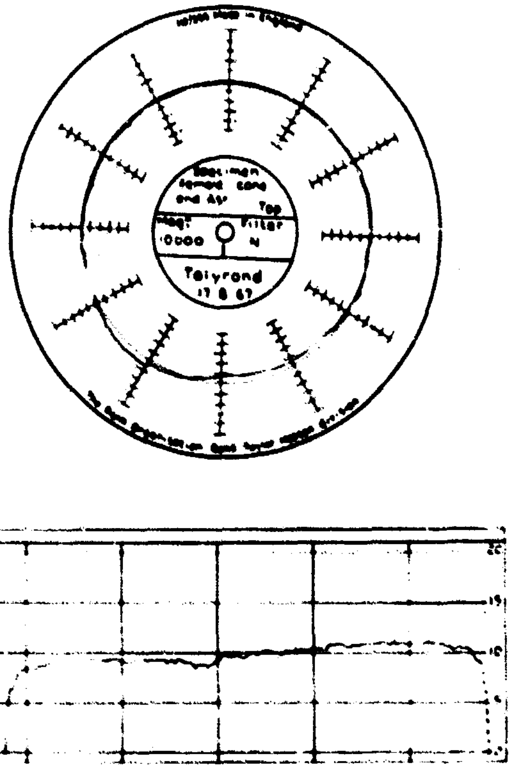
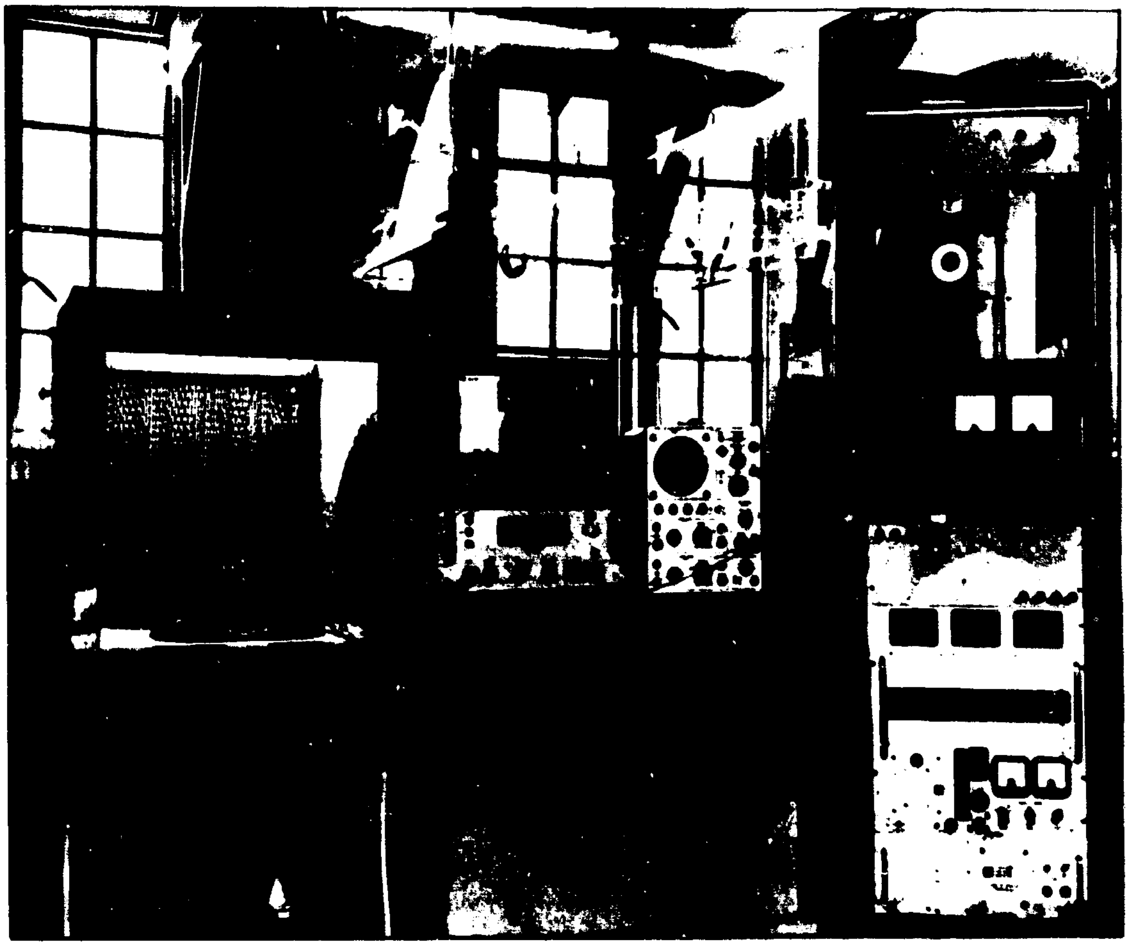


Fig.7 Typical roundness and straightness charts for the second set of bearings.
(In each diagram one small division represents 10 microinches)



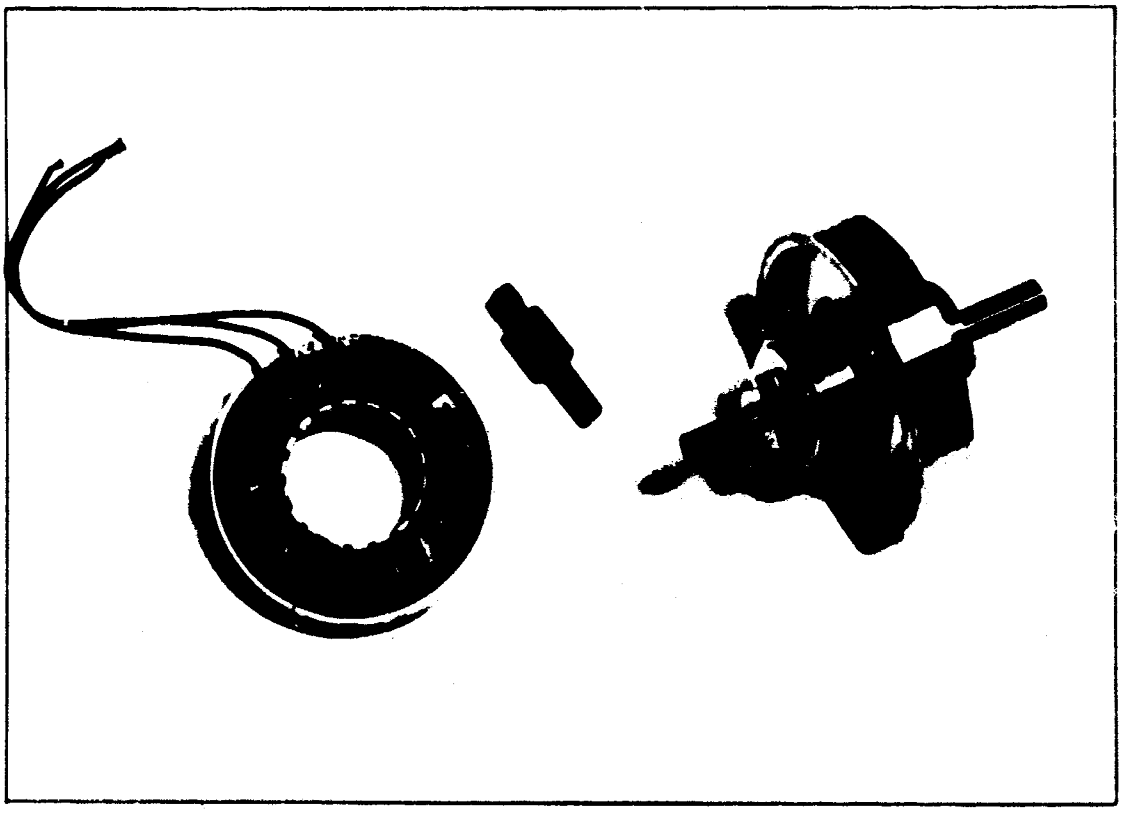
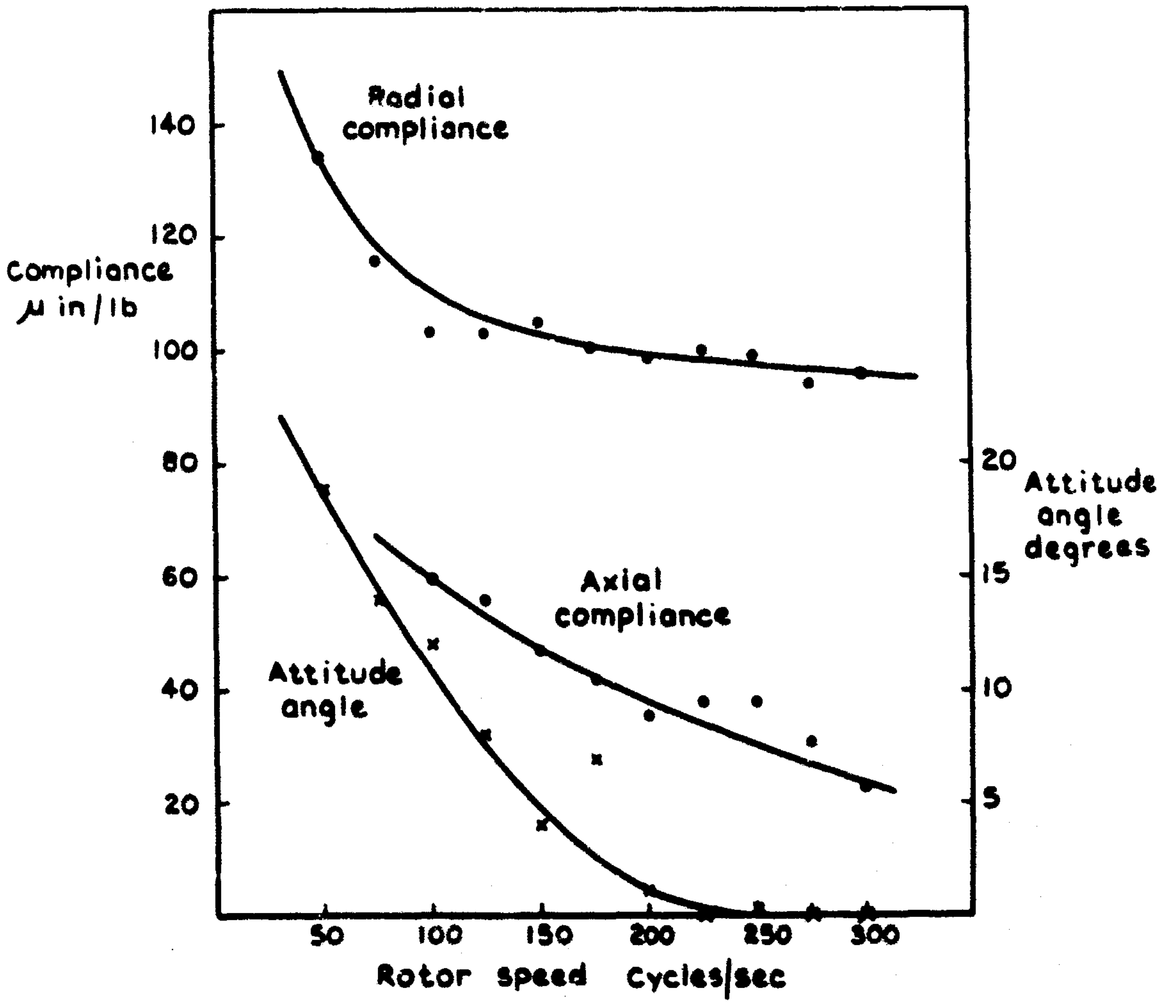
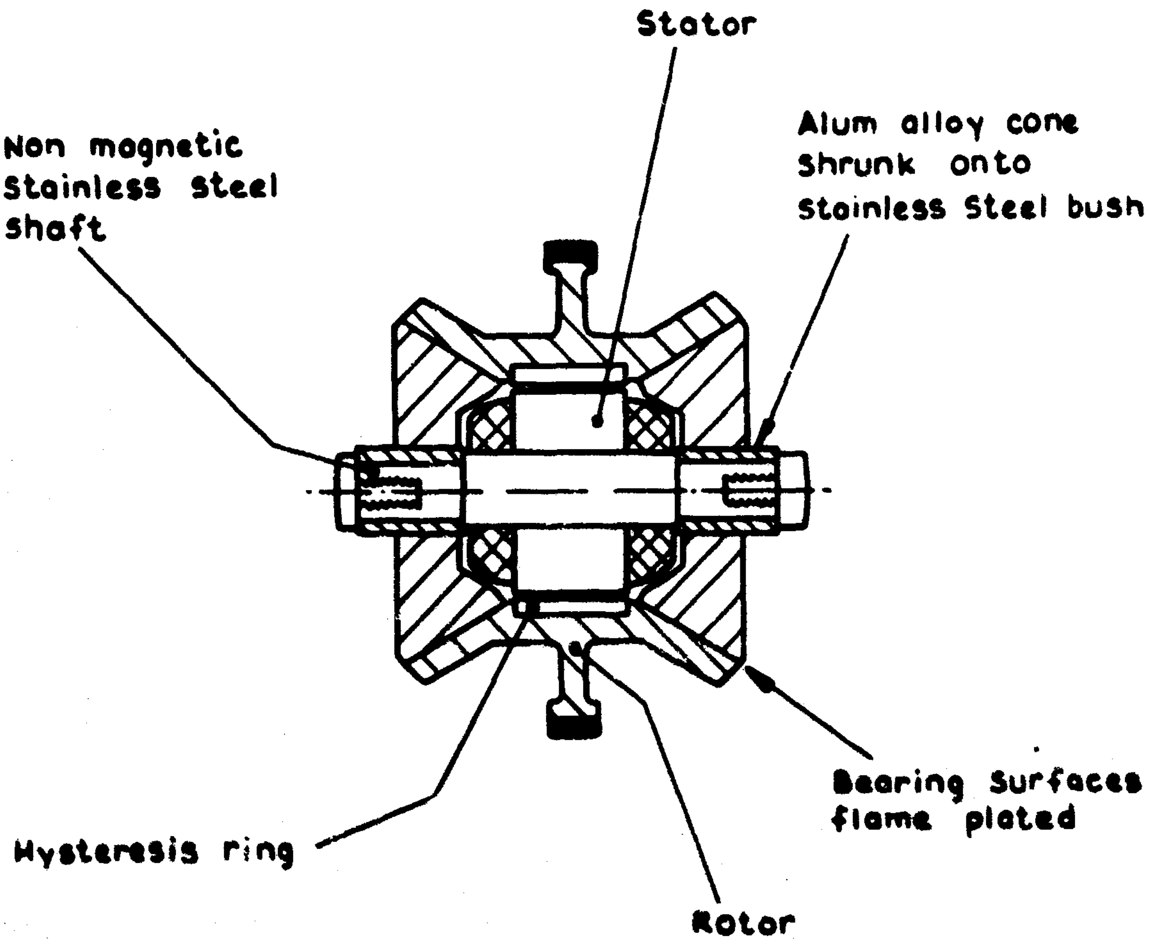


Fig.8 Test equipment and flywheel assembly



Bearing axial length	0·375 in
Grooved length	0·250 in
Groove depth	200μ in
Axial film thickness	200μ in
Groove angle	30°
Land to groove ratio	1

Fig.9 Experimental results for bearing assembly No.2


Stator
Non magnetic Stainless Steel shaft
Alum alloy cone Shrunk onto Stainless Steel bush
Bearing Surfaces flame plated
Hysteresis ring
Rotor

Scale 2:1

Fig. 10 Proposed design for single axis floated gyroscope

STRAPDOWN GUIDANCE COMPONENT RESEARCH

by

Richard J. Hayes

NASA Electronics Research Center
Cambridge, Massachusetts, USA

STRAPDOWN GUIDANCE COMPONENT RESEARCH

Richard J. Hayes

1. INTRODUCTION

The major advantages of strapdown guidance and control systems over conventional gimbaled systems have been cited as reduced weight, volume, power requirements, cost and packaging flexibility. Also, it has been noted that the major disadvantages are associated with the fact that high angular velocities generate performance errors and the strapdown system requires a very fast computer to avoid non-commutative and truncation errors in the attitude computation.

On the basis of these statements, it would appear that strapdown systems hold such significant advantages over gimbaled systems that one would expect to find strapdown systems in wide use today. However, the opposite is true. Gimbaled inertial systems still dominate the guidance and navigation field and probably will continue to do so for several years. Even though the rapid advances in micro-electronics in recent years have made strapdown systems technology more of a reality, many technical problems remain to be solved before the advantages cited above can be fully realized.

This paper reviews the results of research efforts conducted by the NASA Electronics Research Center since 1964 in the field of strapdown guidance and control. The technical areas discussed in the paper include gyroscopes, accelerometers, test methods, computers and electro-optical sensors.

2. INERTIAL COMPONENTS

2.1 Major Differences Between Strapdown and Gimbaled Inertial Techniques

The first practical inertial navigation systems placed the package of inertial sensors, gyros and accelerometers on a platform surrounded by gimbals. The purpose of the gimbals was to isolate the sensor cluster from vehicle angular motion, that is, to keep it rotating in a prescribed manner relative to the stars regardless of the motion of its carrying vehicle. The gyros in this package sense any deviation from the prescribed angular motion and the gimbals are used to return the cluster of sensors to its desired attitude. Accelerometers often perform a dual function in gimbaled inertial navigation systems. In addition to the usual role of using acceleration measurements to compute position and velocity changes, these systems use accelerometer outputs to maintain the sensor package orientation. For example, in a system operating near the earth's surface, accelerometers can be placed with their input axes nominally horizontal. If the inertial sensor cluster is tilted from level, it measures a component of gravity and the accelerometer outputs are used to command gimbal rotations until the instrument platform returns to a horizontal orientation. In these systems, the navigation equations are usually solved in the coordinate axes defined by the nominal orientation of accelerometer input axes and no resolution of accelerometer outputs is required before they can be used by the navigation computer.

In a space stabilized system, where the reference frame is an inertial frame, the gyros see essentially no angular motion. If the coordinate frame is a local vertical frame, it must be commanded by torquing the gyros at the vehicle's latitude and longitude rate. This amounts to, at most, the earth's rotation rate of 15 deg/hr, plus perhaps 25 deg/hr for a transport aircraft.

In strapdown navigation systems, the sensors are mounted directly to the vehicle and sense the vehicle angular motions. The coordinate frame exists in the computer and is calculated from the initial alignment of the system and from the solution of a matrix differential equation using the body axis angular measurements with a suitable integration algorithm. The gyro is now a restrained sensor and the measurement of the angular motion is made by torquing the gyro. The torquing accuracy can deteriorate for very high vehicle rates; however, for the motions of most missions, the torquing must be done with an accuracy of a few parts per million.

The gyros used in gimbaled systems usually have time constants that are much shorter than the significant dynamics of the gimbals, so that they do not present a problem. Gyros in strapdown systems, however, must have higher bandwidth for several reasons. First, for a given output axis inertia, the restraint loop stiffness must be kept high to minimize cross-coupling errors. The gyro error angle should be kept small in the face of large input rates. Second, the gyro must sense as much of the vehicle motion as possible. If the vehicle is coning at frequencies higher than the gyro bandwidth, it will appear as a constant drift error. The accelerometers are compensated for the Coriolis and centrifugal acceleration due to the size effect by using the measured angular motions from the gyros. If the motion cannot be followed by the gyros, then the accelerometer readings contain errors. Lastly, strapdown navigation systems provide body-axis information that could be used for autopilot functions as well as for navigation, for which high bandwidth is desirable.

Strapdown systems promise lower cost, easier maintenance and more reliability since one has the opportunity to achieve redundancy at the sensor level, rather than the system level. On the negative side, removing the angular motion isolation provided by a gimbal structure permits generation of potentially large sensor errors and requires a high dynamic range in the gyros. While single-degree-of-freedom gyros in gimbaled systems are restored to their nominal state by corrective rotation of the sensor package, a rebalance torquing loop is required for each gyro in the strap-down system. This torquing loop introduces several new sources of potential gyro error. In addition, a greatly increased burden is placed on the computer in terms of both speed and size requirements. Calibration of gyro and accelerometer error coefficients cannot be achieved by rotating the gimbals with respect to gravity prior to flight. Therefore, in strapdown systems the stability of the error coefficients becomes of prime importance.

2.2 State-of-the-Art for Inertial Strapdown Sensors

Most existing inertial grade gyroscopes were designed for gimbaled applications and are not directly suitable for strapdown applications. The characteristics of typical existing gimbaled gyroscopes versus those characteristics felt desired for strapdown gyroscopes useful in general space and aeronautical applications are illustrated in Table I. The figures given are maximum values.

As indicated in Table I, many improvements are desirable for strapdown gyroscopes for both aircraft and spacecraft applications. The higher torquing rates are necessary to prevent a loss of reference during situations where vehicle tumbling occurs. The scale factor reduction is necessary for vehicles that maneuver through large angles, i.e., an aircraft taxiing or circling for landing. In addition to those characteristics listed in the table, it is felt that strapdown gyroscopes goals should include the following general requirements:

A capability for in-flight scale factor change.
Gyro failure detection schemes.
Self-calibration of scale factor.

Minimum overall power.
Dual torquer providing compensation.
Life in excess of 50,000 hours.
Over 10,000 start-stop cycles with no degradation in performance.
Rapid warm-up to stated performance in 30 minutes.

TABLE I

Gyro Characteristics	*Gimbaled (Available)*	*Strapdown (Goals)*
Bias stability (deg/hr)	0.10	0.01
Acceleration coefficient stability (deg/hr g)	0.15	0.01
Acceleration squared coefficient stability (deg/hr g^2)	0.04	0.01
Input axis alignment (sec)	7	1
Maximum torquing rate (deg/sec)	120	600
Scale factor linearity (ppm)	200	3
Scale factor stability (ppm)	200	3
Dynamic error linear motions (deg/hr g^2 r.m.s.)	0.04	0.01
Dynamic error angular motions (deg/hr)	9.80	0.01

2.3 Unconventional Strapdown Inertial Sensors

Although it is felt that single-degree-of-freedom gyroscopes with characteristics similar to those listed in the previous section will be useful in many space and aeronautical vehicles, there are other sensors in the research phase which show high promise for some applications. Two of these are the Electrically Supported Vacuum Gyro (ESVG) and the Laser Gyroscope. The main attraction of the ESVG is that it offers a natural inertially fixed reference frame formed by the spin axes of two non-collinear ESVG's. Thus there are no integration errors in the calculation of the attitude transformation matrix and one reads out the vehicle attitude directly. In addition, there are no angular motion errors similar to the cross-coupling or output axis inertial coupling of the restrained sensors. However, the ESVG has serious limitations in its ability to survive in a severe launch vehicle environment, a limited readout accuracy, and undesirable torques caused by its electrical suspension system. The Laser Gyroscope is inherently a more rugged device and, while not a free gyro, it does not have cross-coupling errors and output axis inertia errors. The integration error remains, but is minimized due to the high bandwidth of the instrument. At low rates, mode locking occurs and prevents accurate detection. Additional problems occur in bias stability and linearity. However, using either mechanical dithering or Faraday cells, the dead zone problem has been drastically reduced, with further improvements possible. Some of the linearity problems have been simultaneously reduced along with the reduction of the dead zone. Research is still needed to extend the life of the device to make it competitive with conventional single-degree-of-freedom gyroscopes. The Laser Gyroscope has the most potential for an accurate, low-cost sensor with the least operational constraints, offering quick warm-up, inherently digital output, easily defined and maintained input axis, as well as ease of maintenance.

2.4 Figure-of-Merit Scheme for Strapdown Sensors

During the past several years, the NASA Electronics Research Center has undertaken analytical studies to develop models of strapdown gyroscopes in order to evaluate their potential performance and better understand their error sources. During the course of these studies, a figure-of-merit scheme has been developed for relating the performance of gyroscopes to the accuracy of strapdown inertial reference systems[2]. This figure of merit is independent of the transformation computation and thereby provides a generalized measure of the effectiveness of strapdown gyroscopes.

The figure of merit is the rate of growth of the error in the coordinate transformation matrix (i.e., the drift rate of the computed inertial reference frame) assuming a perfect navigation computer. This can be related to the upper bound on navigation errors arising from system misorientation. The usefulness of this figure of merit comes from the fact that it is possible to express its growth in terms related solely to gyro output errors through a set of equations that are easily implemented on a computer. Consequently, the figure of merit is a useful device for systematic evaluation of different gyros. Because it is unrelated to the direction cosine matrix algorithm, that portion of the strapdown system need not be simulated when comparing angular motion sensors. When a model of error generation within the gyros is provided, the figure-of-merit growth equations permit evaluation of system errors for a given mission to any accuracy desired. Both rate and rate integrating gyros can be treated. Any detail of gyro operation can be included.

As an analysis tool, the figure of merit is used in conjunction with mathematical models of gyro dynamics and gyro error generation mechanisms. With the mission-dependent motion environment as the basic input, the analysis proceeds through calculation of the error histories of the angular motion sensor cluster. The error histories are applied to compute strapdown system attitude errors to any accuracy desired through a set of non-linear difference (or differential) equations. A simple calculation converts the attitude error matrix to the scalar figure of merit. Random motion environments can be treated using Monte Carlo techniques. Of course, the gyro-induced system errors summarized by the figure of merit are highly mission-dependent and any choice of the best gyro made on the basis of this analysis is tailored to the environment described.

In addition to the detailed computer examination made possible by the figure of merit, simplification of the growth equations permits considerable insight into the major effects of gyro errors on strapdown system misorientation. Specifically, it demonstrates the importance of correlations between errors in individual gyros and between errors from pairs of gyros. As a consequence of this insight, gyro synthesis and optimization procedures are possible. Using simplified models of gyro dynamics when necessary, the gyro-induced system errors (as reflected in the figure of merit) can be expressed as an analytic function of the major gyro parameters. Direct determination of the optimum gyro is then possible. Parameter constraints can be easily imposed. Simplifications have been used to permit hand calculations for the solution of simple design problems.

Most importantly, the figure of merit and the equations relating it to gyro errors provide a powerful strapdown gyro design tool that can be used with any degree of accuracy desired.

2.3 Optimum Strapdown Sensor Configurations

The design of an inertial sensor package for gimbaled systems has generally been limited by the need for small, balanced configurations in which the accelerometer and gyro input axes are aligned with the navigational reference frame. In strapped-down navigation systems, the increased computation capability and inherent flexibility in packaging make it possible to design the sensor package for maximum accuracy. For example, in space missions, where thrust is primarily along a single axis, some advantage can be gained through the use of non-orthogonal instrument orientations.

Using a criterion of minimum mean-squared error at the end of navigation, a general computer program has been developed at NASA/ERC for automatically optimizing the orientations to the sensors relative to the body axes[3]. The method of computation is illustrated in Figure 1 and the results obtained to date for two typical space missions, using various available inertial instruments, are given in Figure 2. On the basis of these results, a possible 20-25% improvement in performance seems typical for unmixed systems. The study also shows that the improvement is almost entirely due to a reduction in the effects of gyro errors, and that the optimal gyro configuration is to some extent independent of the actual instruments and mission. The optimum configuration shown in Figure 3 is representative of the results for all cases investigated on the deep space probe. As compared to the standard orthogonal configuration (Fig. 4), the optimum design is remarkably different, with all output axes (OA) nearly parallel to the roll axis and all input axes (IA) close to the yaw-pitch plane (within ±10°). In this configuration, the error is almost entirely due to gyro bias, with each bias uncertainty contribution 6 ft/sec to the midcourse velocity correction.

The curious result of Figure 3 can be attributed largely to the fact that the nominal specific force is always along the vehicle's roll axis. In this case, it is clear that the individual gyro errors due to mass unbalance and anisoelasticity are eliminated by placing the gyro output axes along the roll axis. It is perhaps not so clear that it is also desirable to have all gyro input axes close to the yaw-pitch plane. The reason for this is that the propagation of gyro errors into position and velocity errors depends only on the components of angular error about axes perpendicular to the nominal specific force.

It thus pays to share angular information about yaw and pitch at the expense of angular information about the roll or thrust axis. The degree to which this is done in any specific case depends on a secondary effect in which the roll indication error propagates into an uncertainty yaw, and thereby into position and velocity errors, in the presence of a nominal angular velocity about the pitch axis. The extreme configuration of Figure 3 is thus due to the fact that the total pitch maneuver in the two transfer trajectories considered was only 10-15°; and it is to be expected that the optimum gyro configuration for a boost trajectory would require a more accurate indication of roll.

2.6 Required Sensor Technology Development Areas

An important research area is the angular-motion-induced errors of the strapdown gyroscope, including the dynamic errors and the scale factor error. A better definition of the angular vibration environment of typical vehicles is required. This includes information over a wide frequency range and also on the correlation of motions between body axes. Information about the motion in various locations of the vehicle is required also. To estimate the dynamic errors, an analytical model of the response of the nonlinear pulse rebalanced strapdown gyroscope has been the subject of NASA/ERC analytic studies. Methods of reducing dynamic errors include selection of optimum gyroscope parameters, compensation of the rebalance loop with electronic circuits, compensation by use of angular acceleration sensors or use of the output axis suspension to generate angular acceleration information.

The scale factor error requires study of each element of the gyro and its rebalance loops. The choice of which type of loop, i.e., two-level, three-state or pulse width modulated, is dependent on both dynamic consideration as well as which gives better bias stability, scale factor stability and non-linearity.

The gyro torquer design involves optimal use of geometry and investigation of materials for magnetic, thermal and structural properties. The torquer should have stable magnetic properties with ease of manufacture and a simple stabilization procedure as important considerations. Gyro torquing was recognized as a problem in high performance gimbaled gyros because of the resulting thermal unbalance. Strapdown gyroscopes are especially sensitive to this error because of the high angular rates the gyro sees. This presents a good argument for a constant power restraint loop and also requires the torquer to be designed for the maximum torque with minimum power as a constraint.

Use of gas lubricated bearings for the support of the rotor of a strapdown gyroscope is prompted by the reduction in mechanical noise and the absence of wear between the moving surfaces. The bearings must provide enough torsional stiffness such that the rotor does not touch the stationary portion of the bearing when the instrument is subjected to angular velocities. In addition, the amount of power required to drive the wheel should be kept at a minimum and the design should include low anisoelastic torques. The speed of the wheel should be kept high in order to get maximum angular momentum, which gives lower drift, and the output axis inertia can be kept small, minimizing the angular motion error due to output axis coupling.

Attention has been given to the method of suspending the gyroscope about the input axis. Subjecting a single-degree-of-freedom gyroscope to angular rates about its output axis causes the gyro to precess about its input axis. With no restraint against this motion, rubbing can occur and the subsequent friction about the output axis can be a significant gyro error. Methods of restraint include wires, high-frequency fluid pumping in the vicinity of the output axis pivot, and the electromagnetic tuned suspension systems.

One of the advantages of strapdown systems, the ability to replace a failed sensor easily, leads to the important development areas of mounting and alignment techniques. Since the sensor cannot be aligned in the vehicle by suitable orientations to a known input, techniques are being developed to assure that gyroscopes can be prealigned and normalized. Experiments are underway to evaluate the stability and repeatability of a mounting scheme using a sleeve that fits onto a post having several stages with increasingly tighter fits encountered as the instrument is mounted.

An important error source that is not generally considered involves the rotor being uncoupled rotationally when the case is subject to oscillatory angular motions about the spin axis. The parameters of the motor rotor coupling such as stiffness and damping, are critical in determining this error source. How these parameters are affected by the motor parameters, such as magnetic field strength, saturation levels and geometry are being studied.

The gyro bias error arises mainly from the flex leads that carry the motor current through the damping, and from magnetic interactions between the gyro motor, microsyns, torquer and any magnetic materials in the gyro. By careful attention to materials and their treatment, as well as by good design, these errors can be reduced. The acceleration-sensitive errors come in part from fluid torques, so that thermal design is an important part of the total gyro design. Material dimensional stability and microcreep characteristics are important in minimizing this error source as well. The acceleration squared errors are due mainly to anisoelasticity in the bearing and this is an important part of the bearing design.

An important part of gyro-error coefficient stability is a statistical description of coefficient behavior. With good statistical models of the coefficients, the calibration problem is made easier, since a measurement of gyro drift in a vehicle can be assigned to a particular error source with some degree of confidence.

In addition to improving the performance of gyroscopes for strapdown applications, the method of testing inertial instruments has also received attention. Some of the recent advances in this field are discussed in the next section.

3. INERTIAL STRAPDOWN TESTING TECHNIQUES

3.1 General

Strapdown inertial system testing is a complex task requiring thorough analysis of test objectives, test techniques, test equipment and data processing methods.

A strapdown system test must include both sensor assembly and flight computer, and progressive tests must be made from the sensor level to complete system environmental levels. General categorization of these tests would be

static component calibration,
dynamic component calibration,
system level component calibration,
transformation computer checks,
static system tests,
dynamic system tests,
system environmental tests.

The complete system test is basically a calibration of the coefficients used in the analytical model. This involves experimental evaluation of the random variations of the model's coefficients.

Not only is the system performance demonstrated but also the evaluation is made of each known error source and its contribution to the total system error. System errors must be evaluated ultimately with the hardware in system configuration, even though components have been tested individually, in order to determine the effects of structural, thermal, and electrical coupling on the magnitudes and characteristics of errors.

Total system errors are determined by testing in the desired kinematic environment and differencing the system's output of attitude, velocity, and position, with test reference measurements of the same quantities.

The types of restricted environment tests that are conducted to define magnitude and characteristics of system error sources include:

(i) Single Channel Testing - Constant Input - Consisting of a series of known inputs of rate and acceleration to the inertial sensor unit.

(ii) Single Channel Testing - Vibratory Input - Consisting of a series of known vibratory inputs of rate and acceleration to the inertial sensor unit.

(iii) Coupled Sensor Testing - Vibratory Input - Consists of a series of vibratory inputs applied to the sensors and near-perfect computational processing of their outputs. Errors measured in single channel tests must be subtracted from the outputs of this test.

(iv) Computer Program Testing - Verification that the computer program performs as desired by comparing a known input to a known output. A sensor package simulator is employed in order to precisely control the input.

(v) Coupled Sensor and Computer Testing - Vibratory Input - Consisting of input environments generated by an angular saddle shaker and a combined angular/linear shaker. If effects of coning are to be investigated, a multi-axis rate table is required. Input frequency extends through the bandwidth of the sensors.

In addition to the limited environment component and subsystem tests it is necessary to verify that all system errors are recognized and accounted for. This entails a realistic laboratory environment simulating the intended flight environment:

(i) For Airplane Applications - Requires a multi-axis rate table capable of low frequency, large amplitude motion with low frequency angular vibrations superimposed. Mounting a linear shaker on this table would provide further realism.

(ii) For Space Applications - Added capability of the multi-axis table to provide the angular motion simulating spacecraft on interplanetary missions during non-thrusting phases.

(iii) For Boost and Injection - A centrifuge with the sensor package mounted on a counter-rotating table is a reasonable simultation.

The guidance facility of the NASA Electronics Research Center has started development of the multi-axis test table and angular vibration equipment required for system testing. At present, the capability is limited to the single channel testing which can be accomplished with the single-axis rate table equipped with a two-axis fixture for guidance system loads. This system provides an automatically programmed table axis capable of positioning to an accuracy of ±1.0 second of arc with a resolution of 0.36 seconds of arc on receipt of a digital position command originated at the control console or an external computer.

The axis operates as either a position or rate servo on receipt of a command from a d.c. analog voltage. Maximum acceleration capability about this axis is 10 radians per second squared. The table can follow input signals as high in frequency as 100 c/s. Digital position readout by means of 0.0001 degree and 0.1 degree up-down pulses is provided about the table axis for dynamic measurement of guidance system performance and verification of rate drive accuracy.

3.2 A Precision Tilt and Rotational Vibration System

The ground induced noise on conventional gyro test stations, even those mounted on bedrock, causes serious problems in strapdown system and component testing. The oscillations due to the gyro noise affects gimbaled systems but this motion, or dither, does not impair long-term system performance due to the isolation provided by the gimbal structure. However, in strapdown systems, this noise builds up errors due to the sequential computation technique inherent in the systems design. It is therefore important that these noise signals be measured during instrument tests. At present these gyro noise signals cannot be distinguished from signals produced by ground motions. To avoid this at test stations, attention in recent years has been focused on actively stabilized test pads. At NASA/ERC, analytical and experimental studies have been performed which demonstrate the design feasibility of a tilt and rotational isolation system that will maintain a surface at level to within better than 0.1 arc seconds and provide rotational isolation from ground motions and torque disturbances in an urban laboratory[4]. The combined use of a tiltmeter referenced to gravity for low frequency servo-mechanism control and conventional inertia isolation permits design of a rotational isolation system (about a horizontal axis) having a transmissibility less than one at all frequencies. The principal limitations of the design are the stabilities, resolutions and noise outputs of the sensors used for the servo-mechanism control and the structural natural frequency of the servo-mechanism drive and platform.

To demonstrate the feasibility of the two-axis levelling system that is required for strapdown system testing, an experimental single-axis model of the system has been fabricated and is shown in Figure 5. Tests conducted on the experimental system to date indicate a long-term drift of the order of 0.1 arc seconds and short-term drifts due to gyro noise of less than 0.01 arc seconds in the presence of noise up to 10 arc seconds.

The experimental platform also has permitted demonstration of the system's capability to resist large torque disturbances. For example, a 200 pound man sitting on the platform deflected the drive mechanism 3 arc seconds while the resulting maximum deviation of the platform surface from level was less than 0.1 arc seconds. The instrumentation used in the experimental system is shown in Figure 6. The indicated drift data for the experimental platform is shown in Figure 7 along with the data of the ground motion.

4. STRAPDOWN COMPUTER TECHNOLOGY

4.1 Strapdown Computer Requirements versus Gimbaled System Computer Requirements

In the well known gimbaled system of inertial guidance and navigation, a space-stable reference is physically created by the use of the inertial properties of gyros. Mounted on an inner block, gyros maintain a non-rotating reference by driving the isolation gimbals.

The displacement of the frame can be obtained by integrating twice the outputs of the accelerometers which are mounted along with the gyros.

In contrast to this mechanical arrangement, the space-stable reference frames of a strapdown guidance system is created electronically. The gyros and the accelerometers are fixed to the vehicle frame. The instantaneous relation between the vehicle frame and a known inertial frame is maintained in a digital computer by solving a set of mathematical equations using measurements from the gyros. This information is sufficient for the proper interpretation of the accelerometer outputs. The physical arrangement of a typical strapdown system configuration in a vehicle is shown in Figure 8. Finding the instantaneous values of the transformation matrix which relates the vehicle frame to a known inertial frame is the basic problem of the strapdown system and it is primarily a computational problem. Since the computer integrates the angular motion of the vehicle by the addition of discrete data, a large memory capacity and the ability to cycle computations very rapidly is required to prevent large non-commutative errors and other integration errors.

Many parameters can be used to relate the vehicle frame to the reference frames. Three classes of predominant interest are

(i) three Euler angles, representing successive rotations about three body-fixed axes in a specified sequence;

(ii) nine direction cosines, relating each of the accelerometer axes to the three computational axes; and

(iii) four quarterions or Cayley-Klein parameters.

The nine elements comprising the 3×3 orthogonal matrix are the most straightforward method of describing orientation of one axis frame with respect to another. Solution of the difference equations involves only multiplications and additions and it can be used most easily to resolve the acceleration vector. Since only three of the nine cosine numbers are independent parameters, computational errors in the integration of all nine lead to non-orthogonality of the matrix. However, for short-term applications, such as for boost and injection into a parking orbit, no correction is generally necessary.

Figure 9 is a block diagram for the computation of the direction cosine matrix and the resolution of the acceleration vector into the inertial frame. It represents the additional computational requirement for a strapdown system over the conventional gimbaled system. Let C be the 3×3 directional cosine matrix and let Ω be a 3×3 matrix where elements are given by the rotational rates of the vehicle as readout by the gyro; the computational problem of the instantaneous direction cosine matrix involves the integration of the matrix equation.

$$\frac{dC(t)}{dt} = C\Omega\,, \qquad C(0) = I\,.$$

For a given accuracy, trade-off studies using various integration algorithms, step sizes (or rate of integration) have been obtained. For example, the Lunar Excursion Module (LEM) Strapdown System has found that an integration interval of 20 millisec (or 50 iterations/sec) with a second-order Taylor series expansion integration algorithm is sufficient for its application. Figure 10 gives typically the computational requirements in terms of word length, storage capacity and computational rates necessary to perform the navigation, guidance and control function for injecting into a 100 nautical mile parking orbit using a strapdown system. Figure 11 is a more detailed breakdown of the computational requirements. The attitude computation represents essentially the additional requirements imposed on the computer by the strapdown system over the gimbaled system. It dominates all other computations. Note that the reciprocal of the average operation/second is equivalent to a conventional add operation in a digital computer. Normalization factors are used to convert other operations into add operations. For example, 1 multiplication operation is considered to be equivalent to 6 add operations.

4.2 Modular Computer Technology Development

In undertaking a broad research program in strapdown guidance and control at NASA/ERC, the development of a new computer concept has been a major task. It was felt that present day designs of aerospace computers lack the characteristics of long life, flexibility of application and reliability that are required for future missions. To achieve these goals, modularity in design which provides redundancy of the various elements within the computer (on the level of memory, logic, input/output, etc.) appeared to offer the best chance of success. The advantages of such an approach are:

(i) The physical modules perform recognizable and complete functions to permit the computer to be readily expanded or contracted in capability according to the mission.

(ii) With some initial over-capacity, failures can be tolerated before the mission success is affected. Additional failures can still be tolerated; either all computations are performed at a slower rate, or the least important computational tasks are deleted. With these alternatives, the computer in essence slows down rather than failing completely (a feature commonly known as graceful degradation).

NASA/ERC is developing a modular computer as part of a strapdown guidance system package for use in unmanned space missions. The end result of first phase of this program is the "breadboard" modular computer. The breadboard will be integrated with a strapdown IMU (Inertial Measuring Unit) for total system evaluation. It will also be used as a test bed for research activities in hardware technology as well as the development of software techniques to diagnose and to initiate the switching of modules.

Figure 12 is a block diagram of the configuration of the modular computer. A working computer consists of a working memory module, a working control module, a working arithmetic module and a working input/output module. Data flow between units is in 8 bit bytes plus parity.

1. The CAU (Configuration Assignment Unit) controls the switches which interconnect the various modules to produce the necessary working computer or computers. Communication between the CAU and the CU's is accomplished in 32 bit word parallel form.

2. The CU (Control Unit) performs the duties of determining the sequence of operations within the computer formed by itself in conjunction with one memory, one AU and any applicable Input/Output Unit. It contains 3 index registers, a high-speed local data memory and a high-speed local program memory. A block diagram of the Control Unit is given as Figure 13.

3. The Memory Unit, the Arithmetic Unit and the Input/Output Unit are conventional.

The following is a summary of the specific features of the modular computer:

- 3 independent computers for multi-processing.
- 27 configuration for a single computer operation.
- Byte (8 bit) transfer through switches - this is designed to minimize the number of connections between modules.
- Parity check on bytes.
- 36 bit word consisting of 32 information bit (4 bytes) and 4 parity bits.
- Floating point arithmetic hardware.
- Block transfer of data from/to memory to/from Input/Output Unit. Direct memory access or cycle stealing features for I/O transfer.
- High-speed local data memory (128 words) and high-speed local program memory (64 words) - these are not extensions of the main memory. Each local data memory is addressable. One of the reasons of incorporating this local memory is to speed up the Control Unit/Arithmetic Unit operation. It can be implemented using LSI technology.

- 3 address systems - this is keyed to the local memory concept since the addressable memory is only 128 words. It is particularly suited for strapdown computations.
- 3 index registers per control unit.
- Cross-communication between control units through common and addressable registers in configuration assignment unit.
- 28 instructions.
- 1 μs memory, 4k submodules with expansion up to 32k/module
- Sample instruction rate

Floating point add	12 μs
Floating point multiply	32 μs
Logical operation	12 μs.

It should be noted that, in view of the 3 address system, an add instruction in the modular computer is equivalent to the 3 consecutive instructions for the single address system.

- Simulation results indicate about 4.4 ms will be required for the computation of the attitude matrix in the strapdown system using a second-order Runge-Kutta scheme. At 50 times/sec this is about 25% of the computer load.

This computer is being built by United Aircraft Corporation for NASA/ERC and is presently in the breadboard phase[3]. The deliverable breadboard contains two columns and has two 2 μsec, 4k core memories. There will be 530 printed circuit boards in the breadboard, each board containing an average of 22 integrated circuits. The current estimate is that each module will contain the following approximate number of boards:

AU - 54 boards,	1300 gates,
CU - 104 boards,	2300 gates.

Future versions of the modular computer are planned to incorporate LSI technology.

5. OPTICAL SENSORS FOR STRAPDOWN SYSTEMS

5.1 Requirements

Electro-optical devices can augment a strapdown inertial system during space missions by furnishing updating information to the inertial system, thus providing bounds on gyro drift.

In general, it may be said that none of the stellar characteristics (direction, radiative flux, etc.) will change by a measureable amount due to the movements of an interplanetary vehicle. The fact that they are "fixed" makes them appropriate for defining a celestial framework from which optically derived information can be processed to update vehicle position, velocity, and align the spacecraft.

Vehicle attitude during all phases of a mission can be determined by star pattern recognizers or combinations of near body sensors and star trackers. Near body sensors include sun sensors, planet trackers, horizon sensors, and land mark trackers. When these trackers are used to determine position for initial strapdown inertial alignment the accuracy requirements are 1 to 10 seconds of arc. Spacecraft attitude accuracy requirements are 1 to 10 minutes of arc. For special situations such as laser communication, the fine pointing requirements become 0.01 to 1 second of arc.

Figure 14 indicates how auxiliary optical sensors are combined with a three-axis strapdown inertial measurement unit.

In order to make spacecraft alignment and stabilization more reliable, star field sensors and systems which function independently of most of the other spacecraft systems must be developed (Fig.15). Present interplanetary spacecraft employ sun sensors and a Canopus sensor for spacecraft attitude alignment. These sensors must, however, be oriented by pointing the spacecraft, and are thus dependent on the attitude control system. In addition, the spacecraft must be stabilized in order to point near their specific targets for proper system operation. When sun sensors are used, navigation is mandatory before alignment can take place; otherwise, the line of sight to the sun provides little or no information about the inertial pointing direction. All navigation and attitude computation is then performed on the ground after the sensor data have been transmitted to the Earth.

A planet tracker finds potential application in both attitude reference and control systems and as a navigation sensor. The use of stellar data in attitude reference systems is well known. The stars represent targets which are, in effect, infinitely far from the spacecraft and, hence, valuable in the establishment of spacecraft attitude in inertial space, independent of spacecraft location within the solar system. The use of planets for attitude reference is complicated by appreciable motion of the spacecraft relative to the target planet; accurate ephemeris of the planet is required, as well as accurate location data for the spacecraft. Thus, the use of the planet tracker in attitude reference systems is best served in applications where attitude relative to the near planet is required. This may be the situation where communication antennas are to be directed Earthward or when reconnaissance sensors are to be directed at the planetary surface. In this role, the planet tracker provides a precise planet reference attitude and precludes the requirements for accurate planetary ephemeris and spacecraft position data.

The planet tracker, as a sensor for autonomous navigation, provides two useful data sources: the direction to the planetary center, such that star-to-planet center angle measurements may be incorporated into the navigation system, and the range to the planetary center. During the approach phase, this data is sufficient for accurate navigation measurements and for accurate pointing of sensors at the target planet. Analysis has shown that ranging data is not very accurate at large distances from the planet and the tracker, as a ranging device, could not begin to compete with onboard radar ranging techniques. Pointing information at large distances can be made very accurate, however, 1 part in 500 of the total field of view being a reasonable expectation.

The planet tracker application to the terminal guidance problem circumvents many of the problems associated with the use of an Earth-based system. Uncertainty in the astronomical unit (AU) limits the performance of an Earth-based terminal guidance system. A spaceborne sensor provides data relative to the target planet and offers the possibility of improved accuracy in the terminal and approach phases. As more sensing and computing facilities are carried aboard the spacecraft, the communication problems of a deep space mission such as overloading of existing channels, long times of transmissions for narrow bandwidth data communication, and the time lag for the roundtrip signal passage, are alleviated.

The identification of Canopus involves inspection of the Canopus sensor's output and a man in the loop. Such a system could be used for alignment of a strapdown inertial system's computer if the attitude were then transmitted back to the spacecraft. Reliability of the spacecraft alignment system can be increased considerably by eliminating these interdependencies. Star Field sensor systems involve only a discrete radio command or an onboard sequencer to initiate alignment. The remainder of the system operates autonomously, taking a "snapshot" of the field-of-view, identifying the stars within the field. In addition to providing a computer with the instantaneous pointing direction of the spacecraft, such a system can be operated continuously to generate error signals for the attitude stabilization system.

The star pattern recognition process (Fig.16) involves angular-subtense measurements, relative brightness measurements, and counts of stars within the field. While the computer operates on sensor data, a strapdown inertial measurement unit can be used to measure the spacecraft's attitude change. The solution to the alignment problem at the commanded time is then added to the attitude change for final alignment information

In order to provide a basis for sensor design, an effort was initiated which examined the requirements and trade-offs associated with changes in sensor field of view, limiting stellar magnitude, and probability of observation for an arbitrary number of stars. When any two design parameters are fixed, the third is determined uniquely. Subsequently, optical design parameters, detector characteristics, functional constraints, and performance requirements must be traded off to achieve the required field of view, sensitivity, and probability of observation. The technique employs real stars and involves no statistical modeling concerning star distributions. Probability of observation has been determined with absolute confidence by making use of spherical geometry to find the loci of desired pointing directions within the celestial sphere.

The results of the first two phases of the NASA/ERC study indicate a definite need for higher sensitivity and finer resolution than are presently available in photodetectors. It has been shown that stars of nearly 5th magnitude must be detected for unity probability of observation of just one star in a practical field of view of 20°. Figure 17 shows the field of view necessary to observe n stars with a probability of 1.0, given the detector sensitivity. The necessary field of view is much larger than can presently be achieved with reasonable system accuracies (approximately one minute of arc). The third phase, and by far the most critical phase of the study, has shown that by reducing the probability of observation of one star to 0.95, the field of view can be reduced by more than 30%, and the sensitivity can be reduced by a factor of more than 2.5. Research in the electro-optics field is continuing in order to provide accurate, reliable sensors for updating strapdown inertial systems on deep space missions.

6. CONCLUSIONS

This paper has reviewed some of the research progress made in the field of strapdown guidance and control. The field is relatively new and rapid strides have been made in the last few years. At this time, the following conclusions appear warranted:

(i) The major potential advantages of strapdown systems over gimbaled systems involve reduced weight, volume, power, cost, along with increased packaging flexibility, reliability and ease of maintenance.

(ii) The major limitations are associated with the fact that high angular velocities produce performance errors and the system requires a faster computer.

(iii) In the inertial component area, presently available accelerometers appear adequate, but significant advances are required in gyroscopes for strapdown applications.

(iv) From a systems standpoint, the gyroscopes and accelerometers should be considered as sensing packages, i.e., gyro plus loop and accelerometer plus loop. By designing integral units it becomes possible to replace the sensors and maintain strapdown systems with minimum difficulty.

(v) The laser gyro and electrically supported gyro offer potential advantages over the single-degree-of-freedom floated gyro for strapdown applications.

(vi) The figure-of-merit scheme discussed in the paper offers an analytical means of relating gyro performance to the accuracy of the strapdown system and thereby provides a generalized technique to evaluate the effectiveness of strapdown gyros.

(vii) Strapdown systems permit designers to consider other than orthogonal orientations of gyros and analytic studies have shown that in some cases a 20-25% improvement in navigation accuracy can be achieved by using optimum, non-orthogonal sensor configurations.

(viii) There are still many technological problems to be investigated in inertial strapdown systems. These include material stability, torquer design, alignment procedure, thermal balance and calibration procedures.

(ix) Strapdown technology permits a digital computer interface between guidance and control functions which permits systems designers to consider integrated guidance, navigation and control procedures.

(x) Testing of strapdown systems presents a new series of challenges to system engineers, and some encouraging progress has been made in the field of active isolation from the ground environment. Experimental evidence indicates that noise levels of 0.03 arc seconds peak-to-peak can be achieved using available instrumentation.

(xi) In computer technology, the development of modular techniques appears to have significant advantages over present aerospace computer techniques to strapdown applications.

(xii) Solid state, no-moving-part electro-optical sensors can be effectively used with strapdown inertial systems to provide bounds on gyro drift.

Based on the research progress to date, it appears that strapdown inertial systems will be realizable and find many applications in future space missions. Their utility in areas of long flight where extremely high accuracy is not required appear especially desirable. It is not expected that strapdown systems of the future will replace gimbaled platform systems in those applications where extremely high accuracy is required. By continuing an active research program in strapdown component technology, including inertial, computer and electro-optical sensors at NASA/ERC and other interested organizations, it is believed that the full potential of strapdown systems will be achieved and the results will benefit a wide spectrum of future navigators.

REFERENCES

1. Duncan, R.C. — *Inertial Navigation.* Presented at Institute of Navigation Meeting, Washington, DC, June 29, 1967.

2. Gelb, A., Sutherland, A., Jr — *Design of Strapdown Gyroscopes for a Dynamic Environment.* Interim Scientific Report TR-101-2, Contract NAS 12-508, The Analytic Sciences Corporation, Winchester, Mass., January 15, 1968.

3. Rockwell, D. — *Failure Detection Methods and Optimization of Navigation System Reliability.* (Thesis - to be published.)

4. Weinstock, H. — *Design of a Precision Tilt and Vibration Isolation System.* ScD Thesis, Course II, Department of Mechanical Engineering, Massachusetts Institute of Technology, October 1967, NASA TR-R-281.

5. Manoni, L. — *Modular Computer Design for Improved Reliability in Aerospace Vehicle Guidance and Control Systems.* Presented at AGARD Guidance and Control Panel Symposium on Reliability in Aerospace Vehicle Guidance and Control Systems, Paris, France, March 7 and 8, 1967.

6. Kleiman, Louis A., Arehart, Raymond A. — *An Analytical Approach to the Determination of Stellar Fields of View.* NASA TR-R-257, June 1967.

7. Kleiman, Louis A. — *Sensitivity Requirements for Star Field Sensors.* (Accepted for publication in the AIAA Journal of Spacecraft and Rockets.)

8. Duncan, R.C., et al. — *Guidance and Control Technology for Manned Planetary Missions.* Presented at AIAA Meeting, New Orleans, La., March 5-6, 1968.

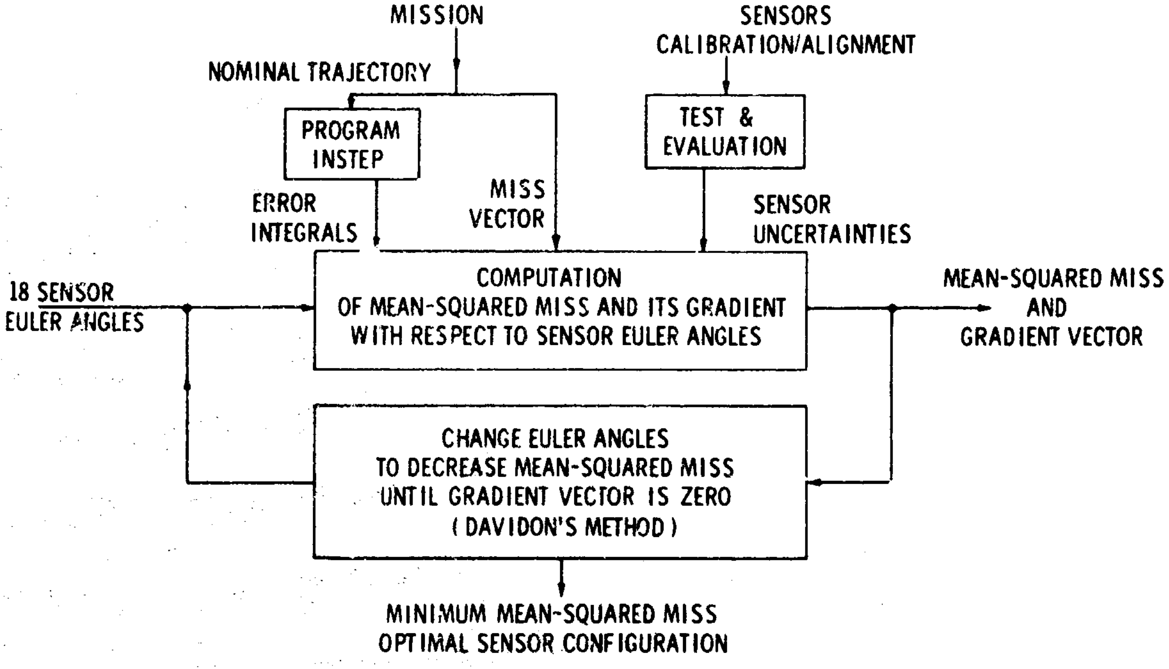

MISSION
NOMINAL TRAJECTORY
SENSORS CALIBRATION/ALIGNMENT
PROGRAM INSTEP
TEST & EVALUATION
ERROR INTEGRALS
MISS VECTOR
SENSOR UNCERTAINTIES
18 SENSOR EULER ANGLES
COMPUTATION OF MEAN-SQUARED MISS AND ITS GRADIENT WITH RESPECT TO SENSOR EULER ANGLES
MEAN-SQUARED MISS AND GRADIENT VECTOR
CHANGE EULER ANGLES TO DECREASE MEAN-SQUARED MISS UNTIL GRADIENT VECTOR IS ZERO (DAVIDON'S METHOD)
MINIMUM MEAN-SQUARED MISS
OPTIMAL SENSOR CONFIGURATION

Fig. 1 Computation of optimal orientations

TRAJECTORY	PERFORMANCE INDEX	STANDARD CONFIGURATION	OPTIMAL CONFIGURATION	PERCENTAGE IMPROVEMENT
EARTH ORBITAL INJECTION	RMS CUT OFF VELOCITY ERROR	6.0 fps	5.4 fps *	9%
		6.0 fps	4.9 fps	18%
DEEP SPACE PROBE	RMS MIDCOURSE VELOCITY ERROR	13.8 fps	10.5 fps	24%
		5.3 fps	4.3 fps	18%
		8.7 fps	4.7 fps	46%

* CONSTRAINED OPTIMIZATION WITH ACCELEROMETER AND GYRO INPUT AXES ORTHOGONAL

Fig. 2 Summary of results. Optimal sensor orientation

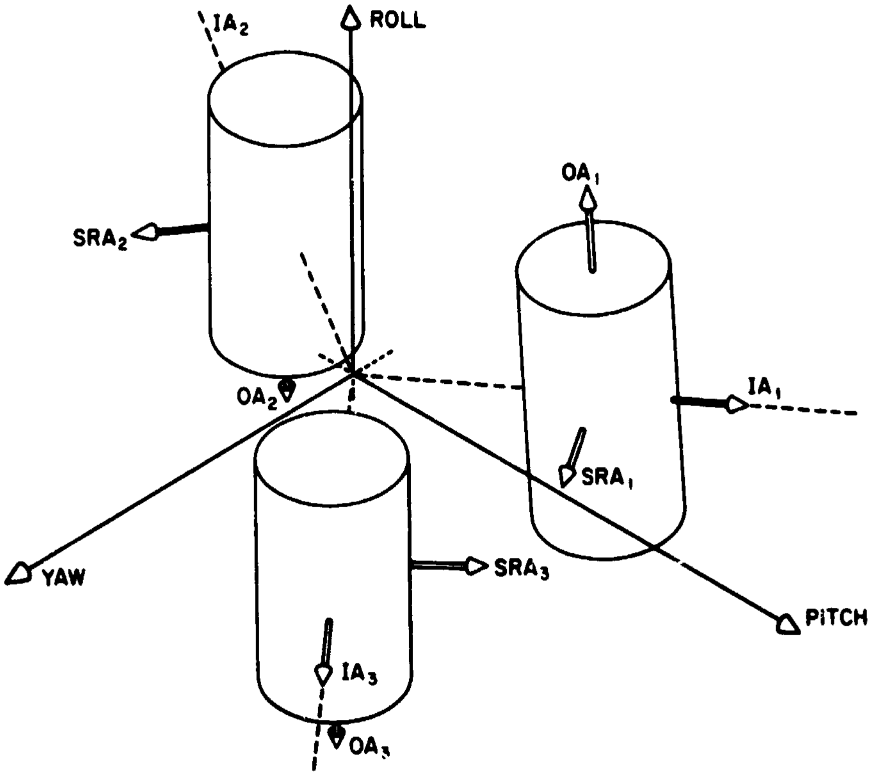


Fig.3 Optimum gyroscope configuration for r.m.s. mid-course velocity correction

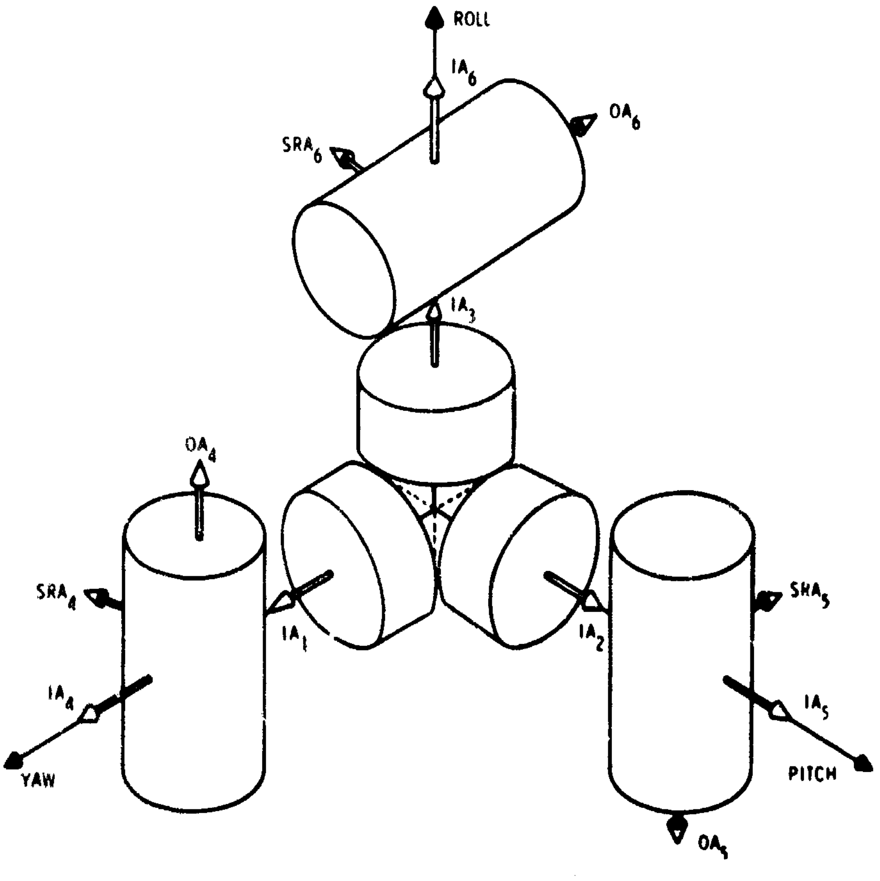


Fig.4 Standard accelerometer and gyro configuration

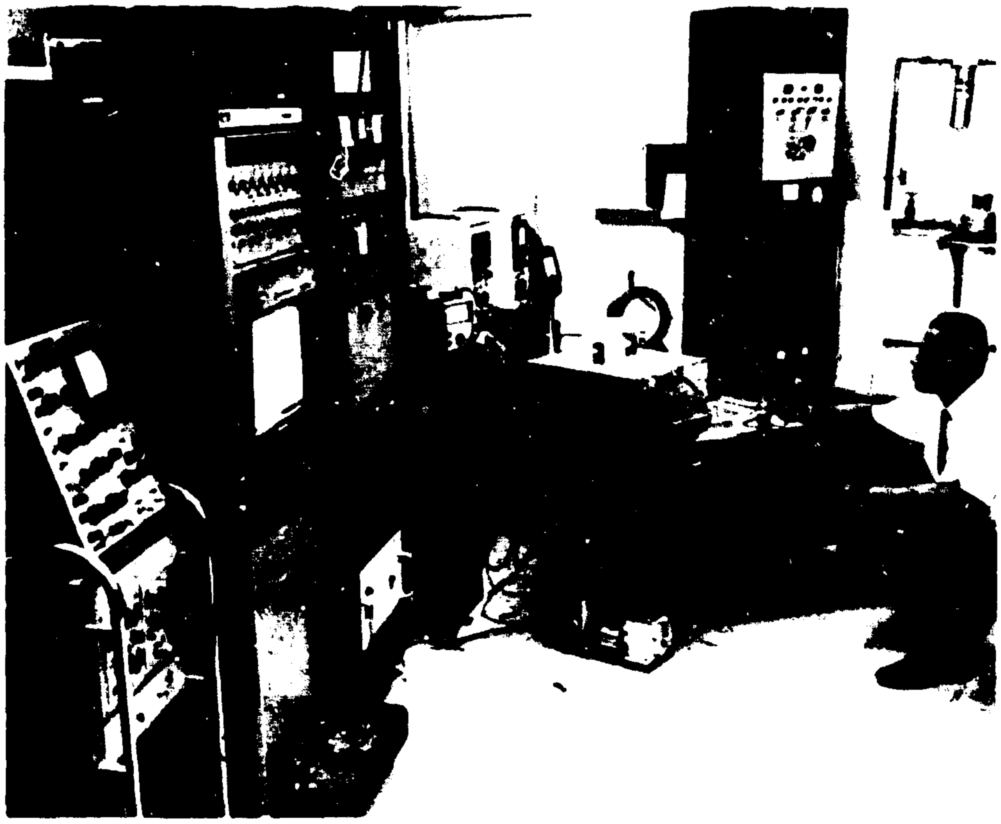


Fig. 5 Experimental tilt station at NASA/ERC

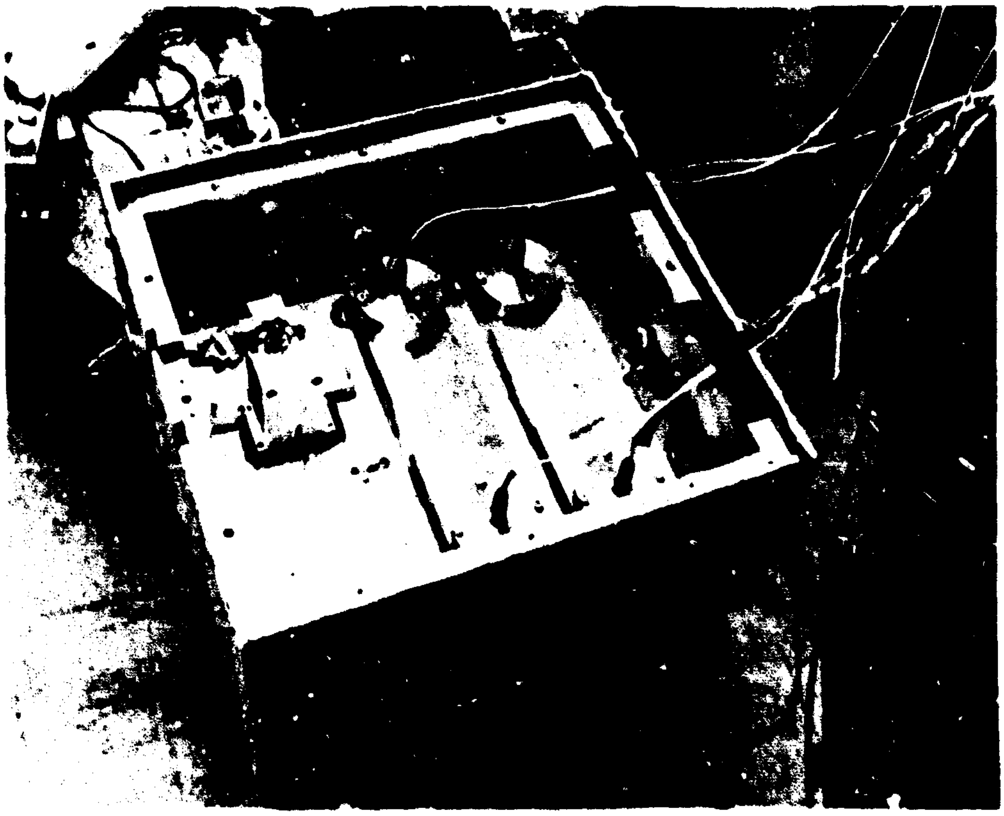
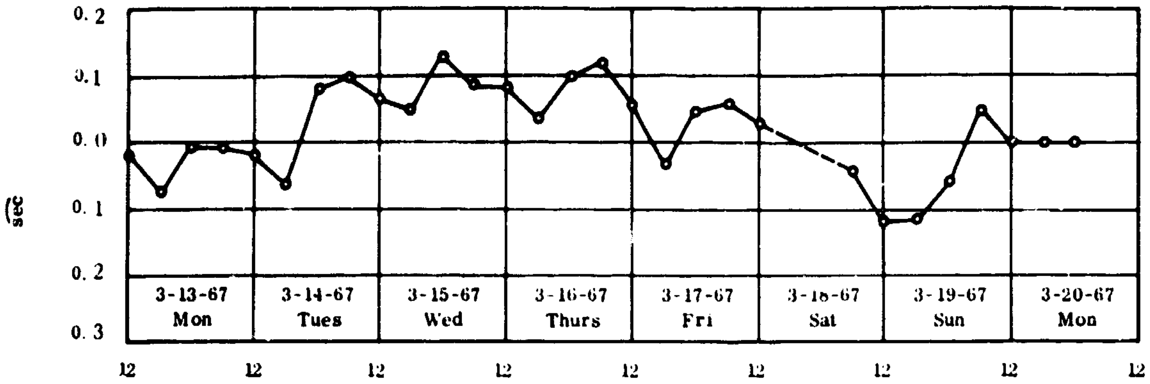
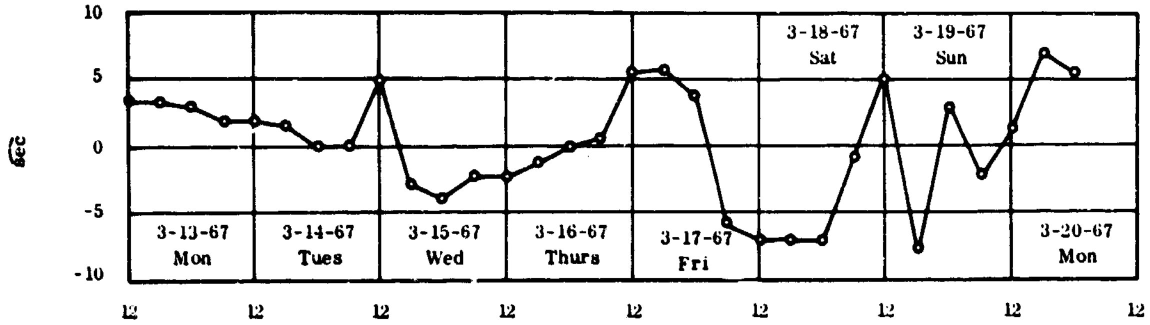


Fig. 6 Close-up view showing control and monitor tiltmeter and control gyroscope



Indicated Platform Error Angle



Ground Motion

Fig. 7 Indicated drift of experimental platform

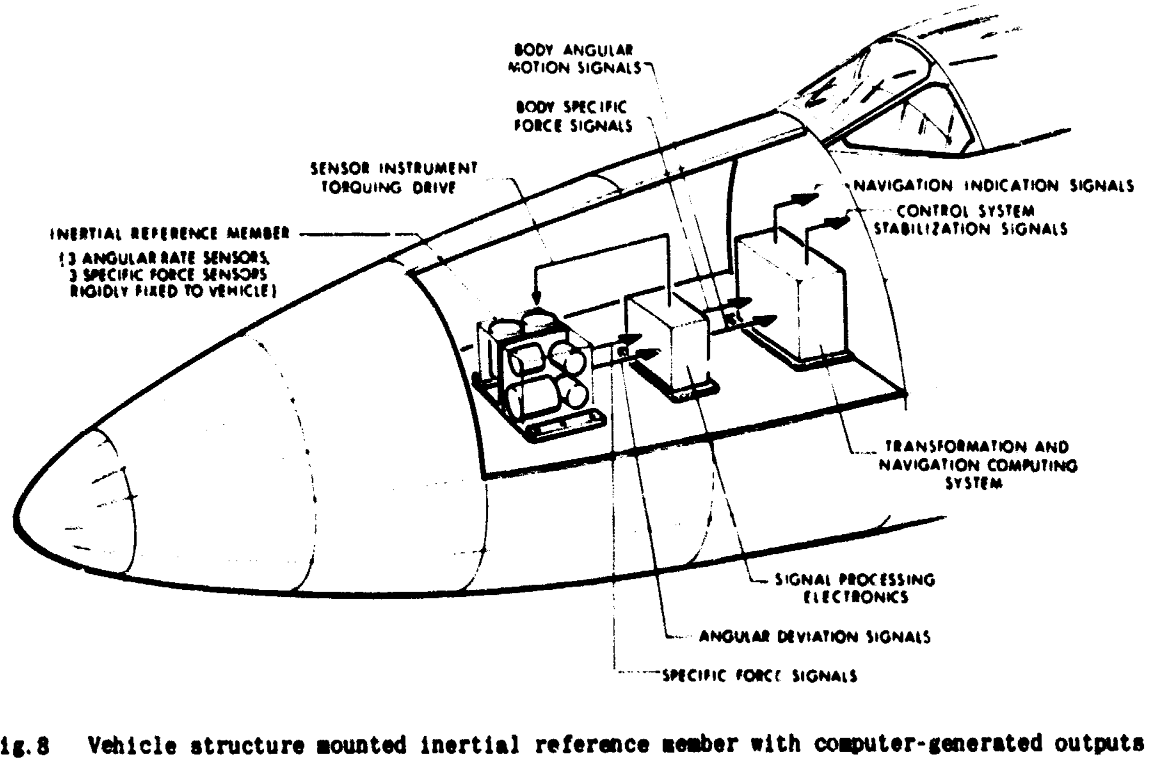

BODY ANGULAR MOTION SIGNALS
BODY SPECIFIC FORCE SIGNALS
SENSOR INSTRUMENT TORQUING DRIVE
INERTIAL REFERENCE MEMBER
(3 ANGULAR RATE SENSORS, 3 SPECIFIC FORCE SENSORS RIGIDLY FIXED TO VEHICLE)
NAVIGATION INDICATION SIGNALS
CONTROL SYSTEM STABILIZATION SIGNALS
TRANSFORMATION AND NAVIGATION COMPUTING SYSTEM
SIGNAL PROCESSING ELECTRONICS
ANGULAR DEVIATION SIGNALS
SPECIFIC FORCE SIGNALS

Fig. 8 Vehicle structure mounted inertial reference member with computer-generated outputs

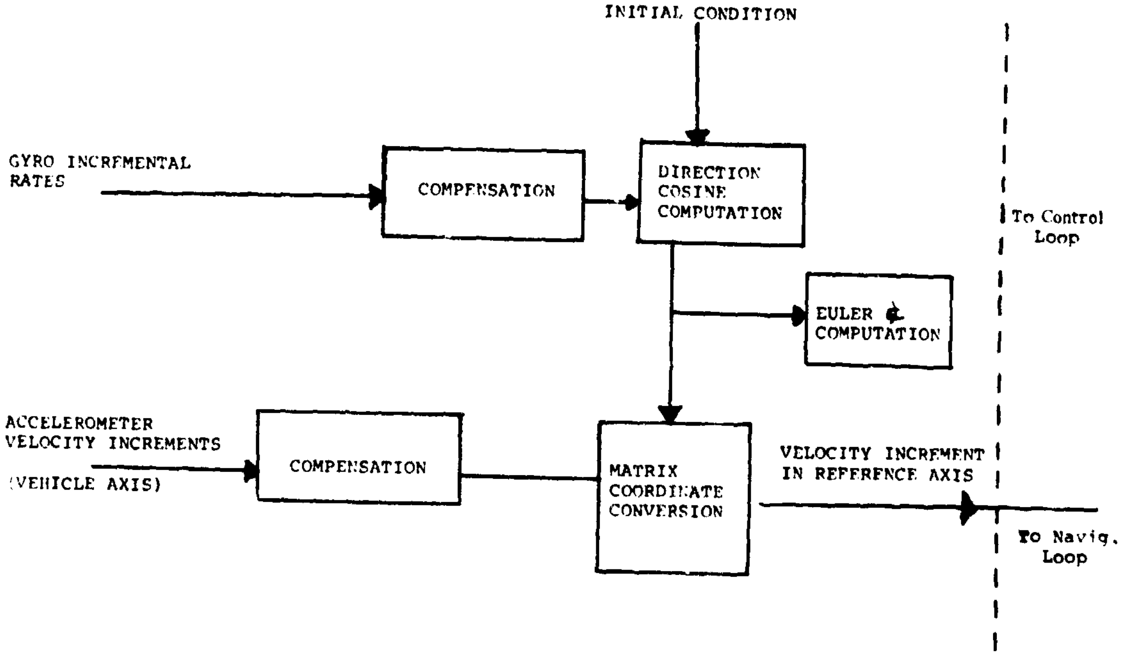

INITIAL CONDITION
GYRO INCREMENTAL RATES
COMPENSATION
DIRECTION COSINE COMPUTATION
To Control Loop
EULER ⊄ COMPUTATION
ACCELEROMETER VELOCITY INCREMENTS (VEHICLE AXIS)
COMPENSATION
MATRIX COORDINATE CONVERSION
VELOCITY INCREMENT IN REFERENCE AXIS
To Navig. Loop

Fig.9 Additional computation in a strapdown system

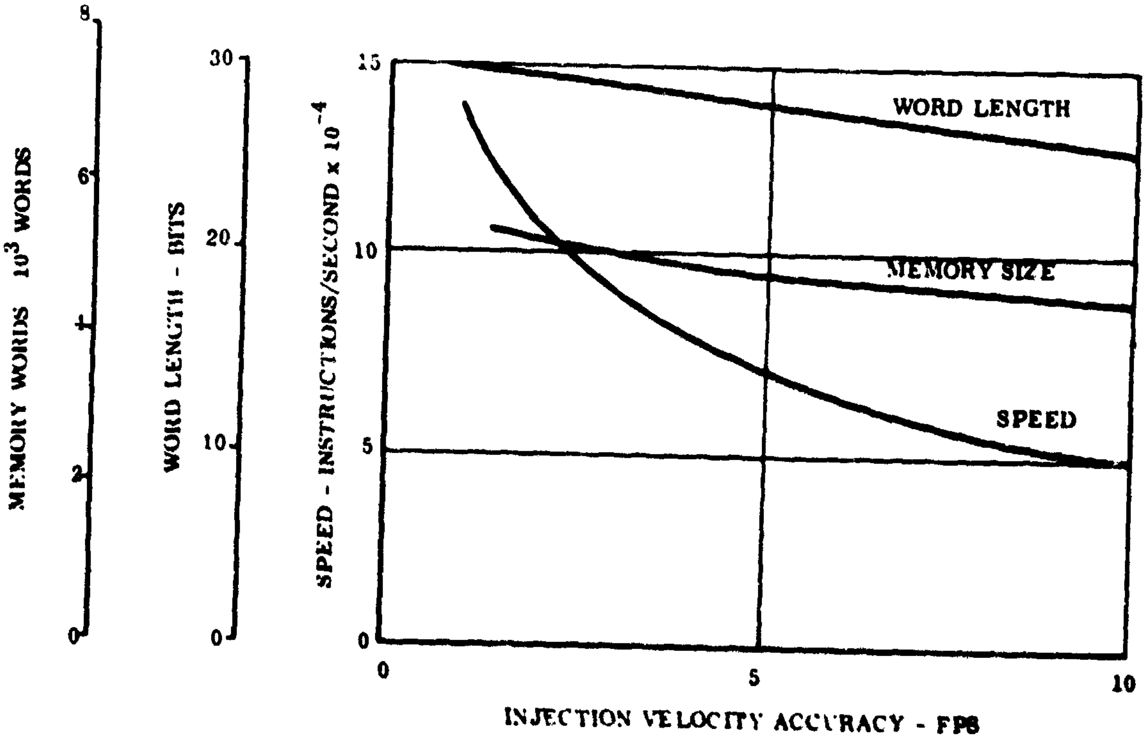


Fig.10 Computational requirements for injection into parking orbit

COMPUTER FUNCTION	10 fps INJECTION ACCURACY			3 fps INJECTION ACCURACY			1 fps INJECTION ACCURACY		
	MEMORY PROGRAM/ DATA	WORD LENGTH* (BITS)	SPEED (INST/ SEC)	MEMORY (PROGRAM/ DATA)	WORD LENGTH* (BITS)	SPEED (INST/ SEC)	MEMORY (PROGRAM/ DATA)	WORD LENGTH* (BITS)	SPEED (INST/ SEC)
ATTITUDE COMPUTATION	435/63	24	28,400	435/63	25	42,600	637/84	28	104,000
POSITION COMPUTATION	482/53	28	4420	482/53	29	5,500	482/53	30	11,100
GUIDANCE COMPUTATION	650/80	21	2600	880/115	22	4,400	1060/135	24	10,600
VEHICLE CONTROL	460/70	16	9200	460/70	16	9,200	460/70	16	9,200

*INCLUDING SIGN POSITION

Fig. 11 Computational requirements for injection into 100 nautical mile parking orbit

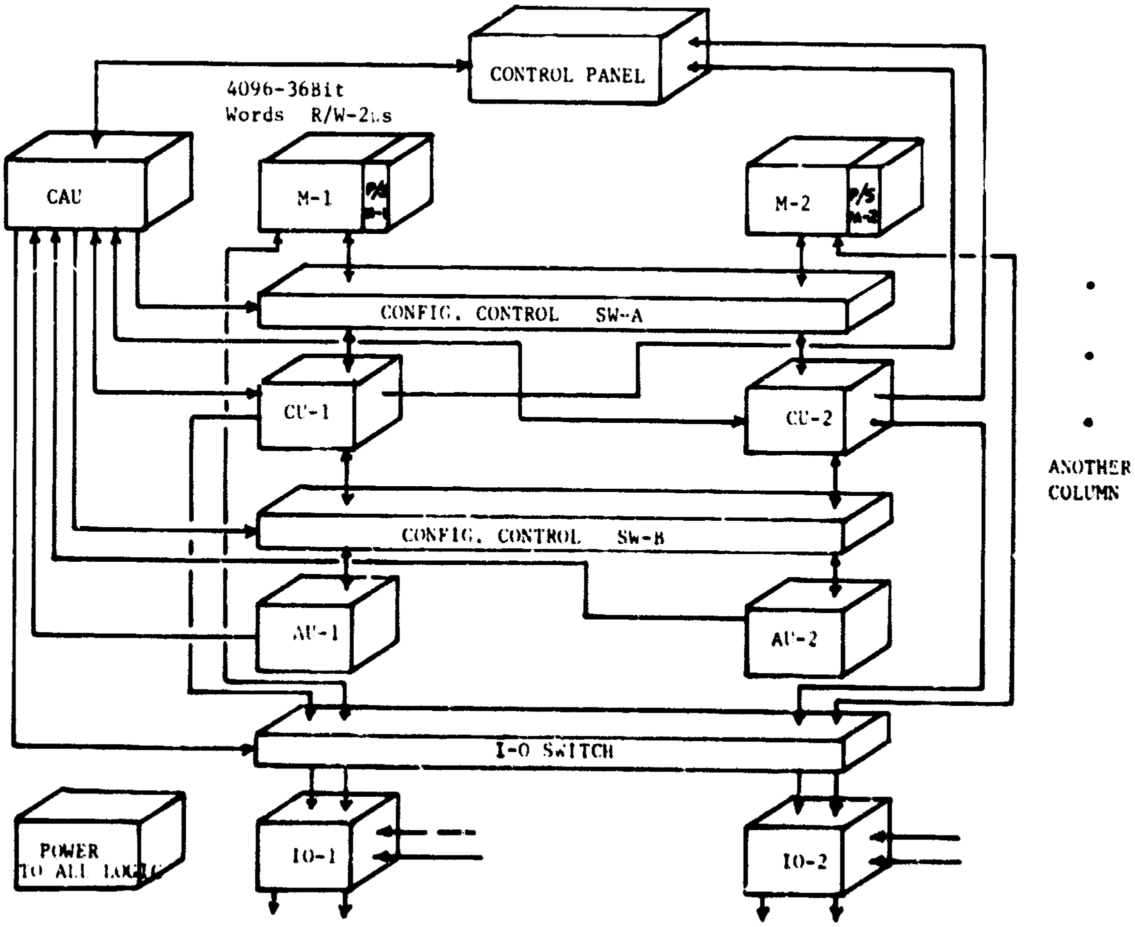


Fig. 12 Configuration of modular computer

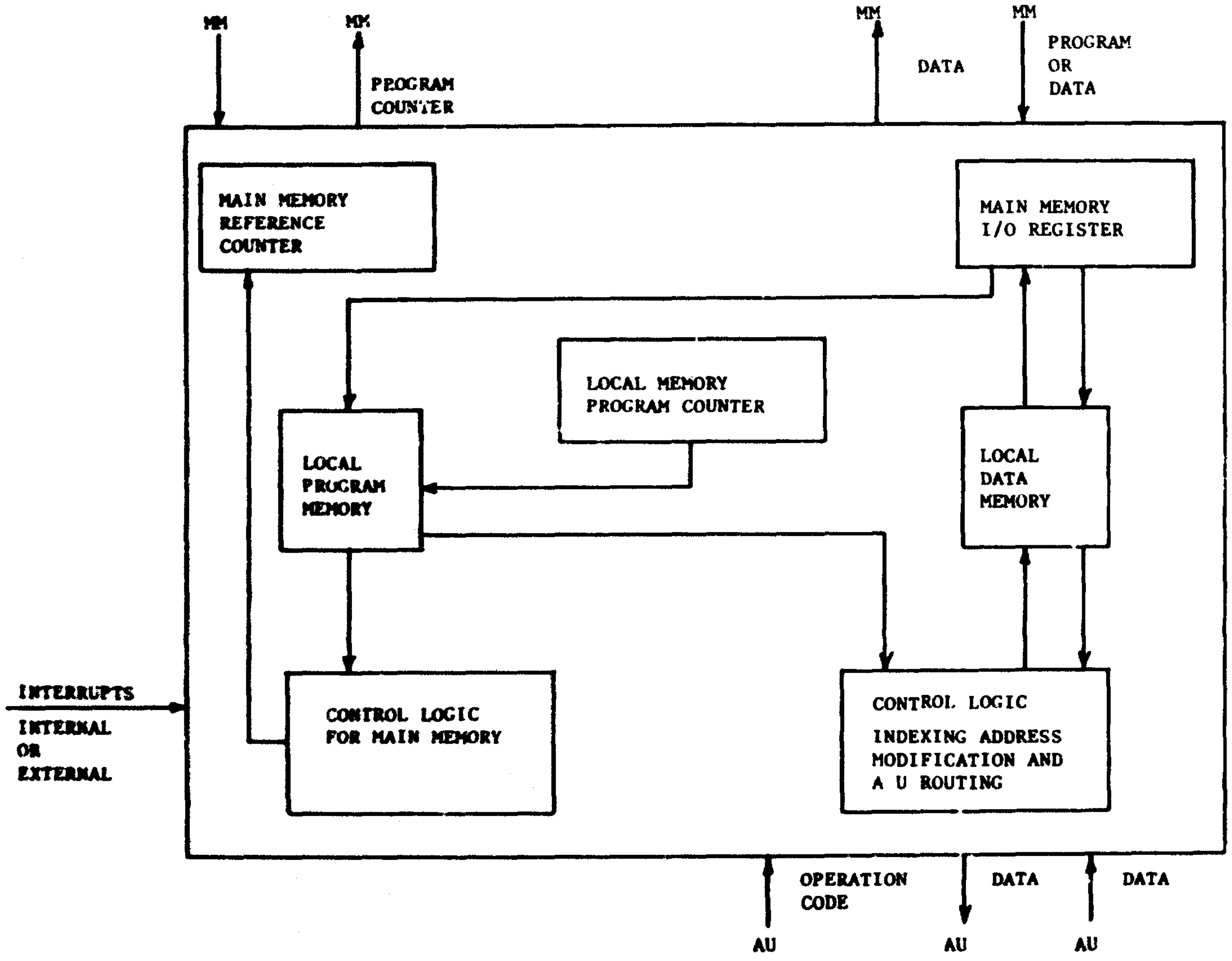


Fig. 13 Modular computer control unit

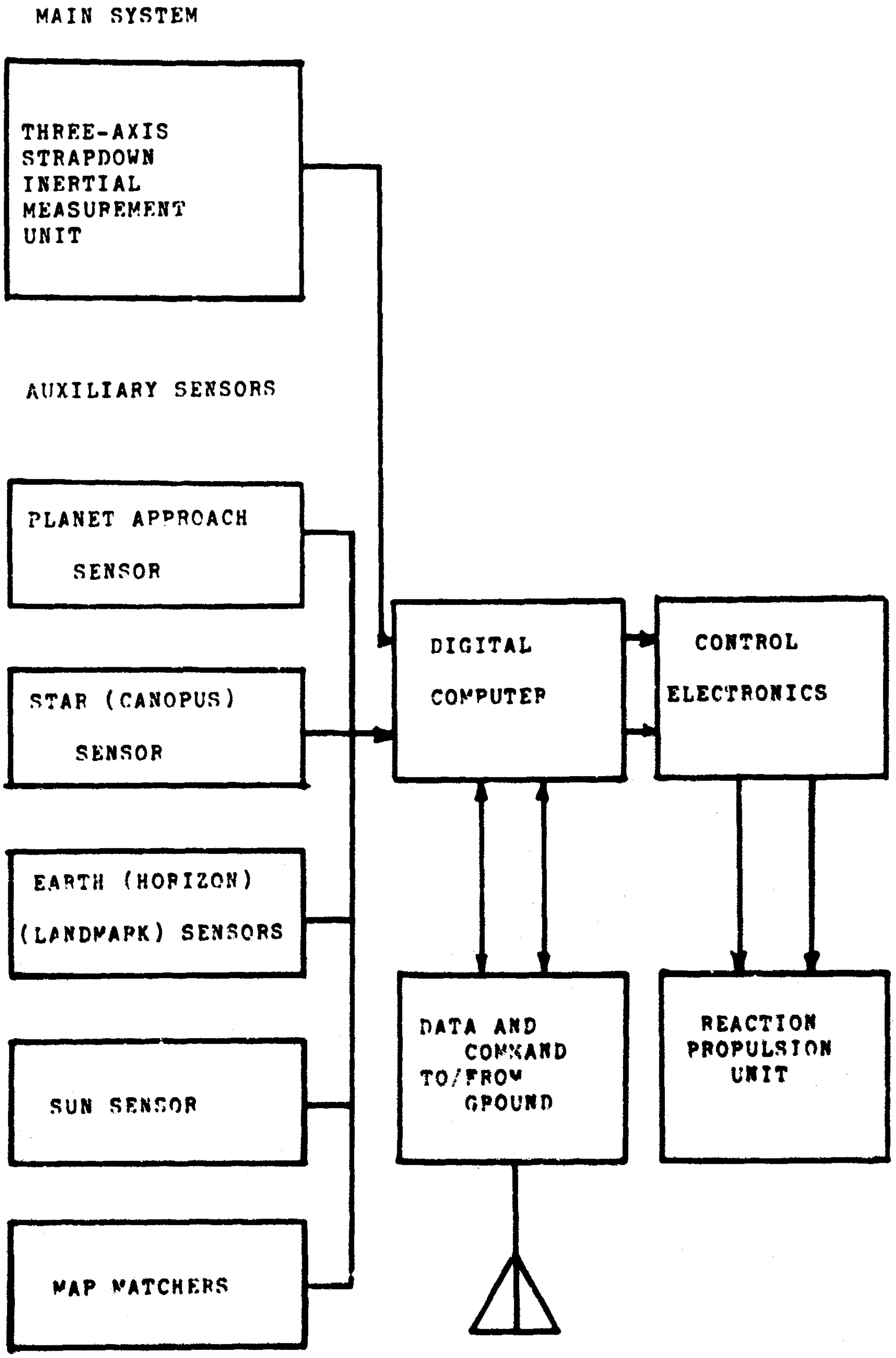


Fig. 14 Spacecraft guidance and control system configuration

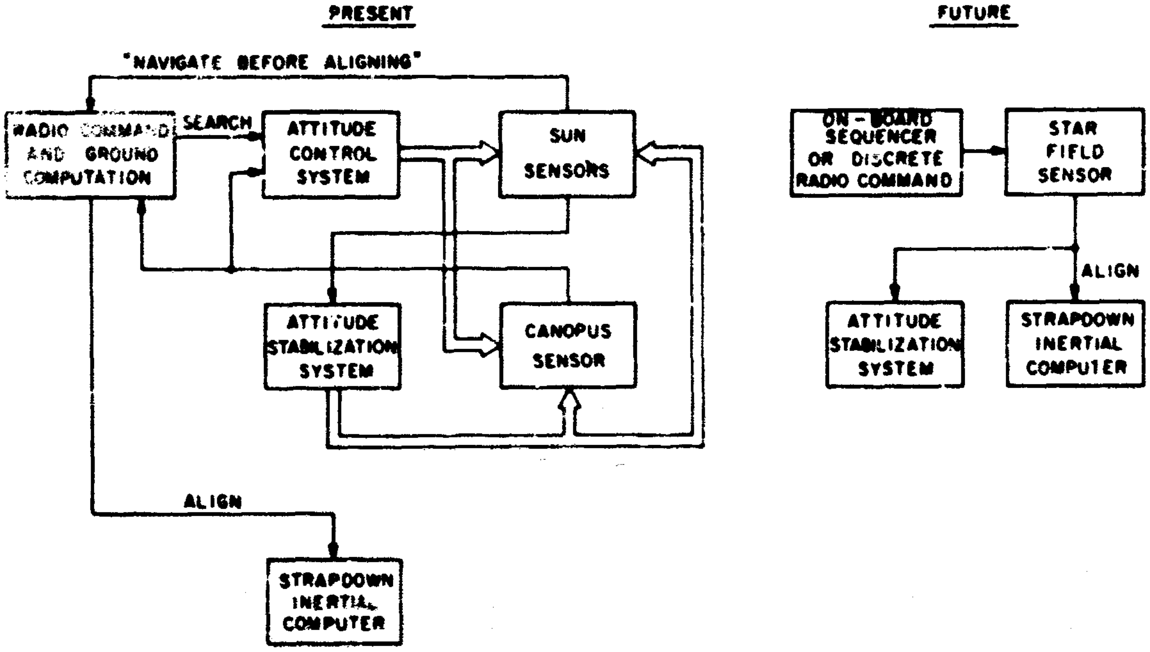

PRESENT
FUTURE
"NAVIGATE BEFORE ALIGNING"
RADIO COMMAND AND GROUND COMPUTATION
SEARCH
ATTITUDE CONTROL SYSTEM
SUN SENSORS
ATTITUDE STABILIZATION SYSTEM
CANOPUS SENSOR
ALIGN
STRAPDOWN INERTIAL COMPUTER
ON-BOARD SEQUENCER OR DISCRETE RADIO COMMAND
STAR FIELD SENSOR
ALIGN
ATTITUDE STABILIZATION SYSTEM
STRAPDOWN INERTIAL COMPUTER

Fig. 15 In-flight alignment

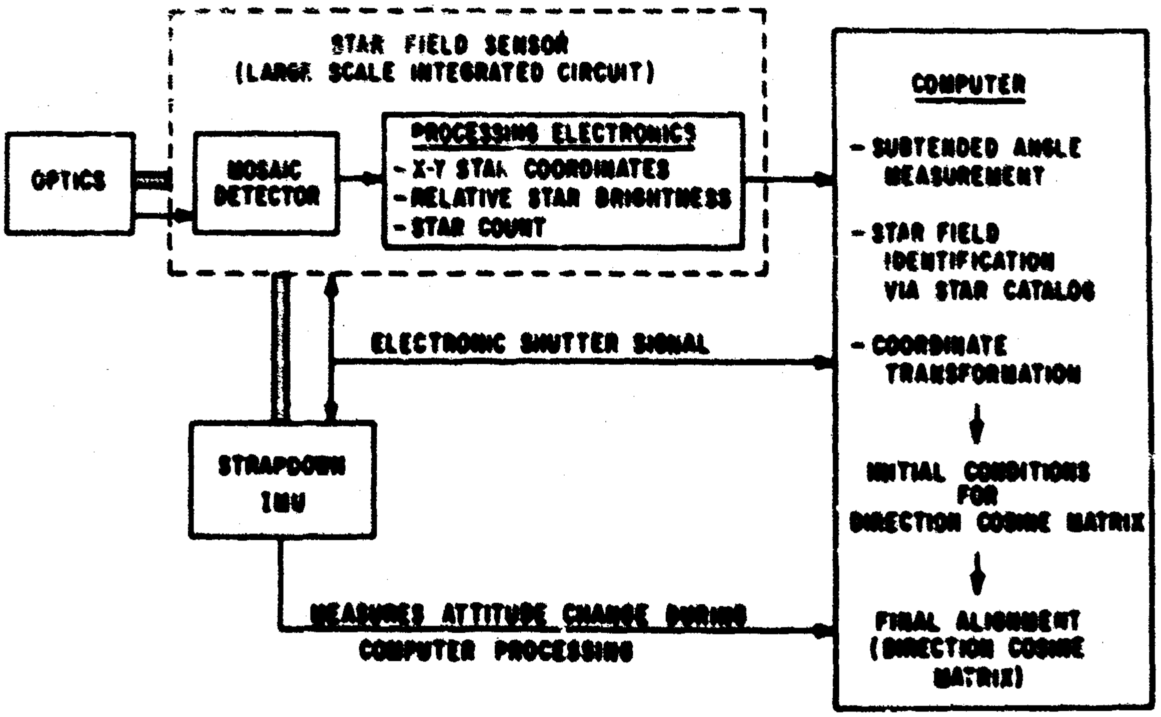

STAR FIELD SENSOR
(LARGE SCALE INTEGRATED CIRCUIT)
OPTICS
MOSAIC DETECTOR
PROCESSING ELECTRONICS
- X-Y STAR COORDINATES
- RELATIVE STAR BRIGHTNESS
- STAR COUNT
COMPUTER
- SUBTENDED ANGLE MEASUREMENT
- STAR FIELD IDENTIFICATION VIA STAR CATALOG
- COORDINATE TRANSFORMATION
INITIAL CONDITIONS FOR DIRECTION COSINE MATRIX
FINAL ALIGNMENT (DIRECTION COSINE MATRIX)
ELECTRONIC SHUTTER SIGNAL
STRAPDOWN IMU
MEASURES ATTITUDE CHANGE DURING COMPUTER PROCESSING

Fig. 16 Optical alignment in space

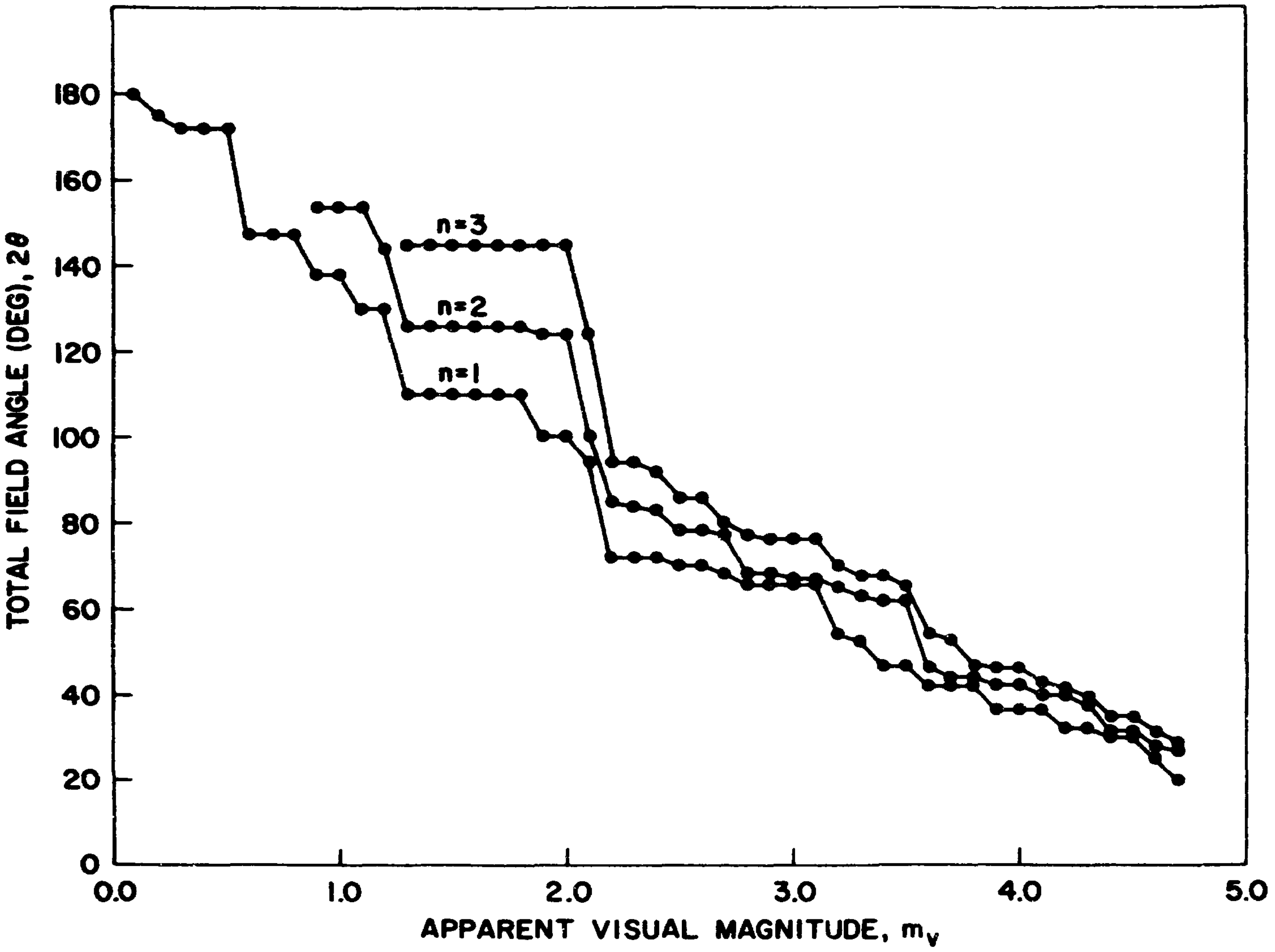


Fig. 17 Field of view required

THE GYROFLEX GYROSCOPE

by

Richard F.Cimera and Michael Napolitano

General Precision Systems Inc.,
Kearfott Group,
Little Falls, New Jersey, USA

SUMMARY

This paper presents a general description of the design and construction of the GYROFLEX Gyro and some of the highlights and results of Kearfott's evaluation testing of units. In the discussion of component evaluation, both bench and test results and environmental test results are presented.

THE GYROFLEX* GYROSCOPE

Richard F.Cimera and Michael A.Napolitano

1. INTRODUCTION

In the early 1960's, industry recognized the need for a small, lightweight, highly accurate but low cost inertial sensor. Kearfott also recognized this need and began a concentrated effort to conclude the development of the Gyroflex gyro. The Gyroflex gyro was conceived in 1958. It is a two-degree-of-freedom, dry, flexure-joint-suspended, free rotor gyro. Its unique feature is a special rotating flexure suspension on one end of a shaft, supporting and serving as a pivot point for the wheel, thereby,providing a support truly free of spherical friction. Internal to the flexure suspension is a gimbaled compensation device which cancels the spring restraint of the flexures. The other end of the shaft supporting the flexure suspension is driven by a synchronous-hysteresis motor. The result is a displacement gyroscope with minimal spring rate operating in accordance with conventional two-axis-gyroscopic theory.

Figure 1 is a mechanical schematic of the Gyroflex gyro. The axis not shown has identical pickoff and torquer coils. The Gyroflex gyro concept permits the inertia wheel to be completely free of all wound components. The armature of the two axes pickoff and the permanent magnet of the d.c. torquer are integral parts of the inertia wheel, but they do not require electrical connections. Therefore, there are no requirements for slip rings, rotary transformers, or flex leads to transmit power through the suspension system. This is an important performance and reliability advantage.

2. SUSPENSION SYSTEM

To achieve inertial quality performance, the suspension system must have very low spring rate and no friction. Flexures provide a suspension system that has no friction because there are no rubbing or sliding parts. Low spring rate is obtained by using the thinnest possible flexures. High strength is required to resist the stresses encountered during shock and vibration. The flexures are fabricated to operate well below the microyield point. Careful control of grain size is necessary because of the thin section. Very low hysteresis is obtained through material selection and precise manufacturing controls.

Since the suspension system rotates with the inertia wheel, the effect of spring biases in the flexures is cancelled for a complete rotation.

The Gyroflex gyro has an additional desirable characteristic - no "dead-band"; the smallest motion of the wheel with respect to the case is sensed by the pickoff.

3. SPIN MOTOR BEARING WEAR AND MASS STABILITY

In the Gyroflex gyroscope, the spin bearings are not located on the inertia wheel assembly. The inertia wheel is separated from the motor by the flexure suspension.

---

*General Precision Systems Trademark

It follows that any shift in the position of the shaft of the motor, either because of preload changes or raceway wear, cannot cause a mass shift of the inertia wheel with respect to the flexure suspension. It merely causes a shift of both wheel and suspension an equal distance. Similarly, shifts in the position of either the spin bearing retainers or spin bearing lubricant do not affect the Gyroflex gyro performance. Such shifts in a conventional gyroscope would directly cause a mass shift along the spin axis.

Another significant reason for the excellent mass stability of the Gyroflex gyro is the complete absence of wound components and potting compounds on the inertia wheel. It is well known that these components and compounds are somewhat unstable and do contribute to mass shift in conventional gyroscopes because of outgassing, with subsequent weight change and migration of materials.

4. FILL MEDIA

Since the flexures support the inertia wheel of the Gyroflex gyro, fluid support is not required. The fill media is dictated by the requirements of the spin bearing lubricant and the thermal time constant. An optimum pressure of gas is utilized to attain long life and to achieve rapid warm-up. Proper design of the inertia wheel and wound components, to allow large clearances, minimizes the effect of the gas.

Other important performance and reliability advantages of the Gyroflex gyroscope are its relative insensitivity to contamination of the gas or the presence of foreign particles, except in the spin bearings, and the absence of fluid stratification which frequently is a problem in floated gyros.

5. THERMAL SENSITIVITY

The sources of thermal sensitivity in the Gyroflex gyro are the same as in conventional gyros, except for those associated with the use of floatation fluid. The Gyroflex gyro does not use floatation fluid; therefore its thermal sensitivity depends mostly upon the thermal characteristics of the metals used in its fabrication. The principal sources of thermal sensitivity are pickoff null shift with temperature changes and mass unbalance shifts along the spin axis with changes in temperature due to mismatch in linear thermal coefficients. These have both been minimized.

The absence of floatation fluid permits fast warm-up. With no fluid thermal gradients, the drift rate is constant as soon as the metal parts reach temperature. The part with the largest time constant is the inertia wheel, because it is suspended on the flexure, but its time constant is only two minutes.

6. TEST DATA

The inertial navigation version (Mod II) of the Gyroflex gyro has been tested at the Naval Air Development Center, Johnsville, Pennsylvania, and Holloman Air Force Base, Alamagordo, New Mexico. The results have been excellent. Kearfott has conducted full bench and environmental testing on all versions of the Gyroflex gyro on component level, and early versions have been flight tested in the LCI system at Holloman AFB and in the GPL Hereflex* system. The component data is included. The LCI system has achieved an average in-flight accuracy of better than 1 nautical mile per hour radial error rate. The Gyroflex gyro in the Hereflex system maintained its original calibration to within 0.02°/hr during 550 hours of operation during the flight test program.

---

*General Precision Systems Trademark

7. RATE MODE TEST METHOD

Gyroflex gyro drift testing is conducted in the rate mode. In this manner both axes can be monitored simultaneously without the use of expensive two-axis servo tables. In the rate mode the signal from the X-pickoff is amplified and demodulated, and applied to the torquer, which provides torque about the Y-axis with such a polarity as to drive the error signal to null. In a similar manner, the Y-pickoff is maintained at null by amplifying its signal, demodulating, and applying current to the torquer which provides torque about the X-axis. These currents fed into the torquers are, therefore, directly proportional to the sum of all the torques about the two axes. By monitoring these currents on a strip chart recorder, an instantaneous presentation of drift rate is obtained (Fig.2). To facilitate random drift calculations and to increase the sensitivity of the data, digital data have been taken. A voltage-to-frequency converter converts the d.c. signal from the sampling resistors to an output frequency which is proportional to the input voltage (f = ke). This frequency is counted by the counter and the preset counter determines the time increment during which the counter counts. When the selected time increment expires, the preset counter commands the printer to print the count accumulated on the counter. This digital readout is proportional to the time integral of instantaneous drift rate for the particular time increment. A typical time for the selected time increment has been 2 minutes and typical times for the length of the run have been 2 and 3 hours (Fig.3).

8. RANDOM DRIFT

Drift is measured in a rate-capture mode over a three-hour period in a 150°F ambient temperature. Drift is recorded both with the spin axis vertical (SAV) and spin axis horizontal (SAH). Random drift is computed by dividing the drift run into two-minute time intervals and averaging each two-minute interval. These averages are then used to calculate a 1σ-value of random drift. Test results are depicted in Figures 4 and 5.

9. DAY-TO-DAY DRIFT

Typical day-to-day performance is depicted in Figure 6. Each data point represents the drift level after shutdown and cool down to +78°F. Prior to this series of tests, the gyro was subjected to a complete environmental test program.

10. SERVO MODE DATA

Typical servo mode random drift and torquer scale factor linearity data are shown in Figure 7.

11. ENVIRONMENTAL TESTING

11.1 Shock

A gyro was subjected to eighteen, 30 g, 11 millisecond, non-operating shocks. Three shocks were applied along each of six axes. Performance before and after shocks repeated within test equipment accuracy.

11.2 Vibration

A gyro was vibrated to the full ASN-57-Gyrocompass Attitude Reference System (GARS) vibration specification, MIL E 5400. Curve 4 applied to the vibration isolator, sweeping from 10-500-10 c/s at a rate of 1 octave/min; the natural frequency of the isolator was 20 c/s. Drift was measured during vibration and was calculated by integrating over time intervals of one minute. Performance before and after vibration repeated within test equipment accuracy.

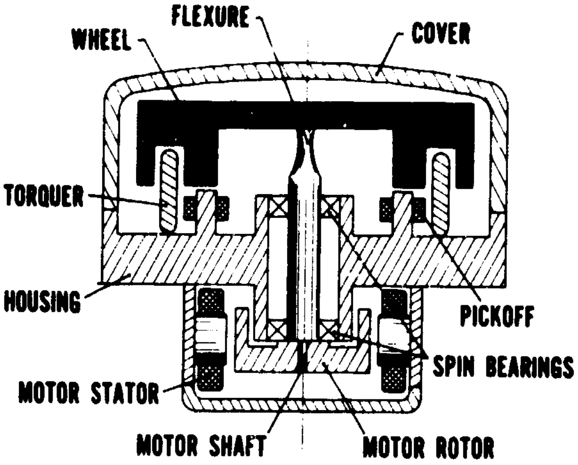


Fig. 1 Gyroflex gyro mechanical schematic

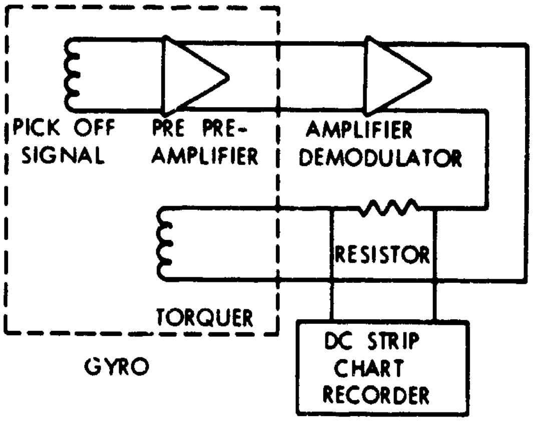


Fig. 2 Rate loop testing (recorder monitoring)

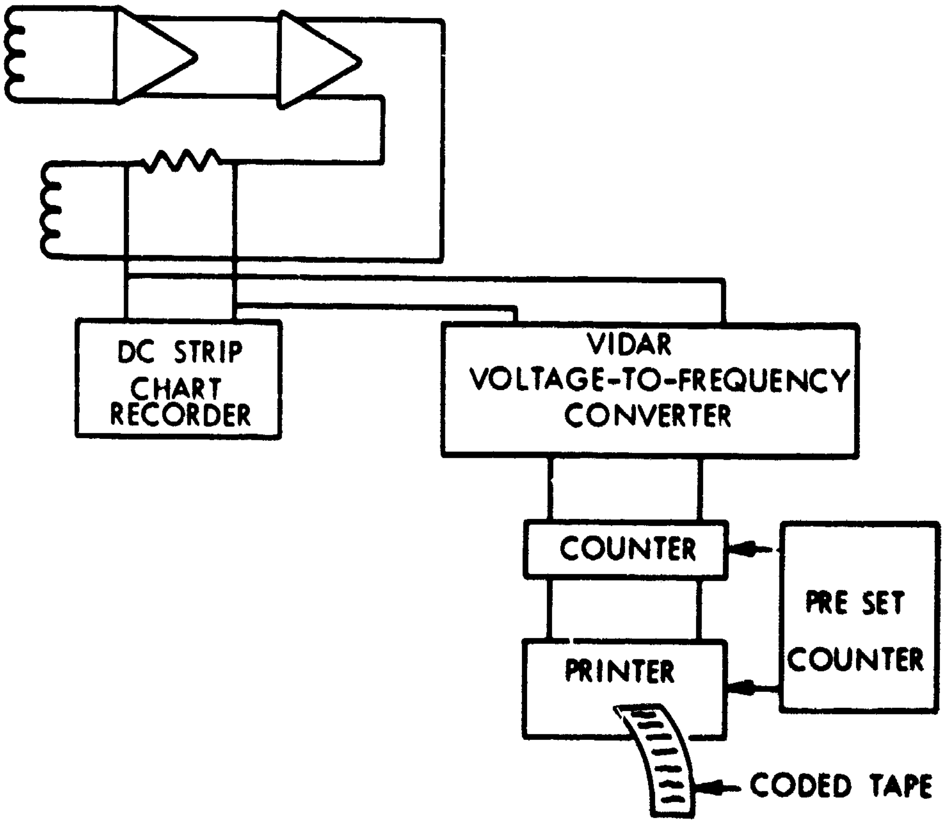


Fig. 3 Rate loop testing (printed data)

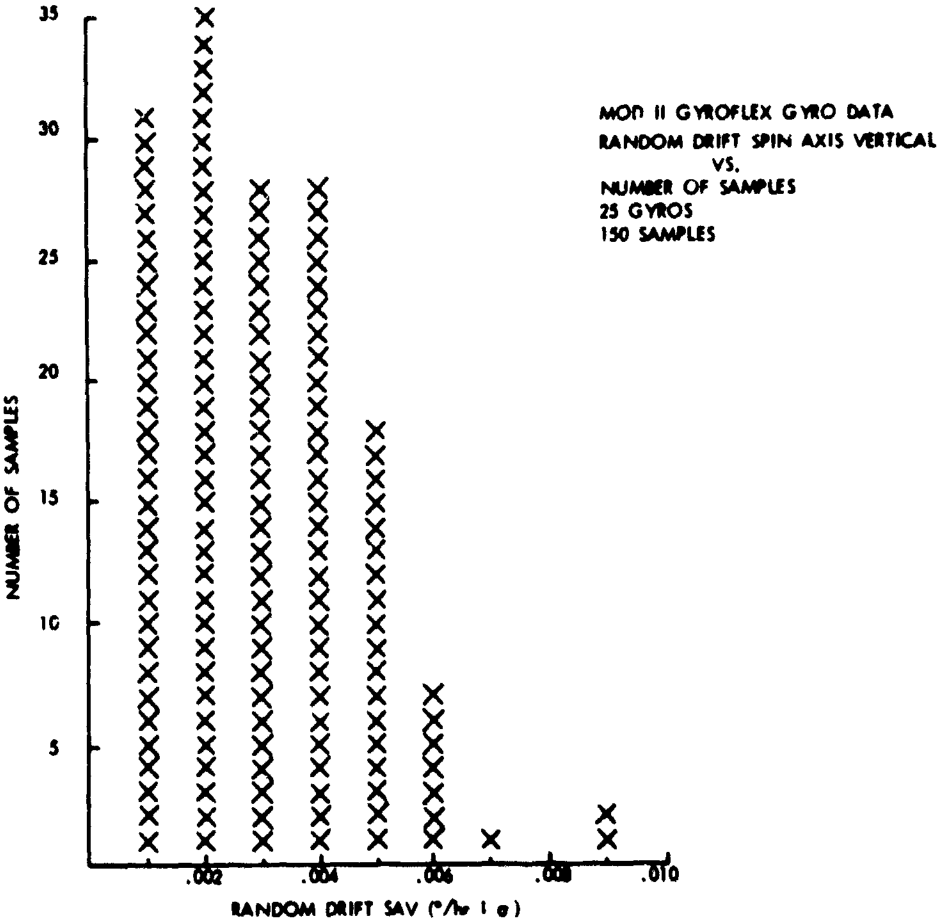


Figure 4

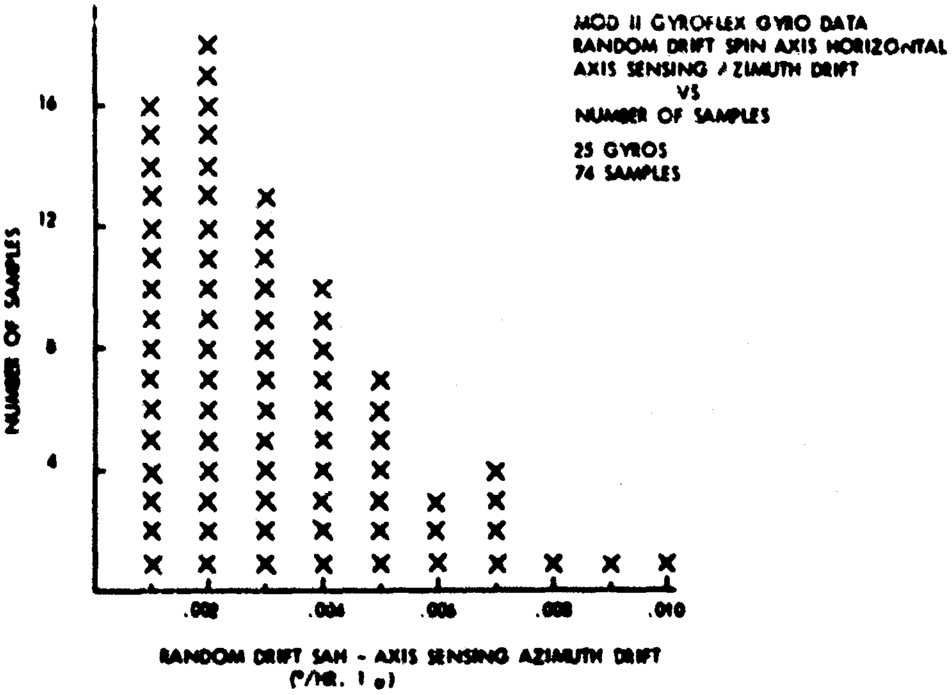


Figure 5

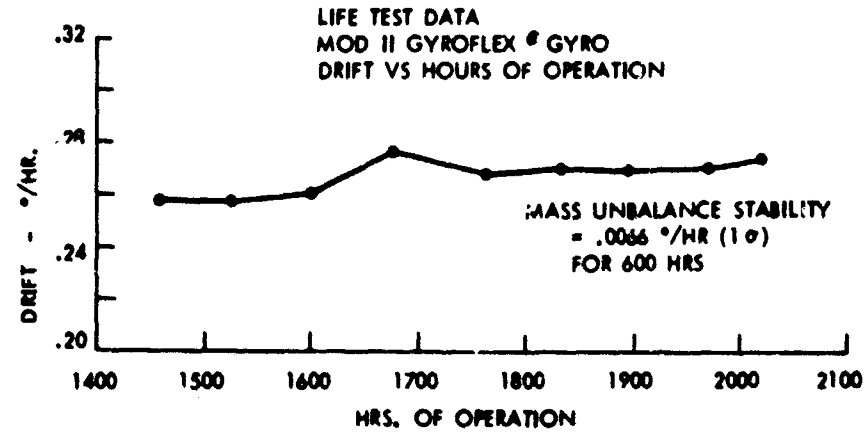


Fig.6 Gyroflex gyro Mod II life test data

FREE - SERVO MODE RANDOM DRIFT

NUMBER OF INTERVALS	LENGTH OF RUN	PITCH AXIS		ROLL AXIS	
		AZIMUTH	LEVEL	AZIMUTH	LEVEL
16	1 1/2 HRS	.0018°/HR (1 σ)	.002°/HR (1σ)	.0032°/HR (1σ)	.0015°/HR (1σ)

SERVO MODE - TORQUER SCALE FACTOR LINEARITY

RANGE OF INPUT CURRENT	PITCH AXIS ERROR	ROLL AXIS ERROR
± .30 MA TO ± .75 MA	.0042% - 1σ	.0042% - 1σ

Fig.7 Test data

THE ROTOR-VIBRAGYRO

by

E. Mühlenfeld

Institut für Schwingungsforschung
74 Tübingen, Stauffenbergstrasse 21
West Germany

THE ROTOR-VIBRAGYRO

E. Mühlenfeld

Rotor-Vibragyro is another word for the rotating, two-axis vibragyro, under development elsewhere under the trademarks "Vibrarotor-Gyroscope" or "Oscillogyro". This gyro is considered as a rate gyro, but the following will be equally valid for a rotor-vibragyro with low damping which, for platform stabilisation purposes, may be regarded as a free gyro.

Figure 1 helps to explain the principle of its operation. A motor is rigidly mounted on the aircraft or platform. One end of its shaft carries a bearing, which allows a resonator to oscillate about an axis perpendicular to the motor axis. Consider a mass positioned above the motor axis and moving backwards with velocity v_1 due to the rotation. If the vehicle rotates about the vertical y-axis, a Coriolis force F_1 is generated, in a direction to the right, perpendicular to both the rate and velocity vectors. When the mass under consideration is below the motor axis, the velocity vector, and consequently the Coriolis force, have changed direction. The resonator is thus excited by a torque which is alternating with rotational frequency ω. The centrifugal force generates a restoring torque. Rotation about the x-axis excites the resonator with a phase shift of 90° and can thus also be measured.

The basic difference between conventional rate gyros and the Rotor-Vibragyro, from the viewpoint of communication engineering, is now illustrated.

If an aircraft or other vehicle is rotating with angular velocity Ω, a gyroscopic reaction moment M acts on a rotor, spinning about an axis, which is forced to follow the motion of the vehicle (Fig. 2). In conventional rate gyros this moment is transferred to the gimbal frame by radial forces, exerted by the rotor shaft on its bearings, thus rotating the gimbal frame about an angle α against the restraining spring. The bearings, being a link of the transmission channel, add their noise to the signal. The need for low-noise bearings involves high manufacturing costs and limits operating life.

The inherent noise has led to attempts to avoid bearings altogether, resulting in the tuning fork and other vibragyros. As the difficulties encountered by vibrating masses on a straight line are numerous, the rotor-vibragyro, like conventional gyros, moves masses on a circular path but removes the indispensable bearings from the transmission channel of the signal. The gyroscopic reaction moment is converted into an angular deflection ϑ of a part of the rotor, called the resonator (Fig. 2). This oscillating deflection relative to the rest of the rotor is converted into a signal, which can be matched to a suitable channel to transfer the signal from the rotor to the aircraft body. An electrical signal is exposed to disturbing fields in the transducer as well as in the transfer channel and thus requires careful shielding. Therefore an optical pick-up was chosen and the pick-up of the first experimental model is shown in Fig. 3. The rotor, the outline of which is given by the broken line, is mounted on the shaft of a motor fixed in the vehicle. The rotor carries three mirrors, one of which is attached to the bar-shaped resonator, and torsional vibrations of this bar about an axis perpendicular to the rotor axis are to be measured. All other parts of Figure 3 are fixed in the vehicle. A square light source is projected onto a pattern, the beam of light being reflected by three mirrors. For the time being we assume the resonator to be fixed within the rotor. Due to the reflection at the rotating, slanting mirror, the square image will move round the rotor axis in the plane of rotation, as long as the resonator is in the beam. A deflection of the resonator will now shift the path of the square image either up or down. By passing through the inclined slot of the pattern, the image causes the photocell behind the pattern to produce an electrical pulse, which occurs at a time depending on the inclination of the resonator.

We want to be sure of measuring the inclination of the resonator relative to the rotor and do not want the signal to be affected by displacements of the rotor axis within its bearings. A similar pick-up on the left side of Figure 3, with a sensing mirror attached to the rotor instead of the resonator, produces electrical pulses which are shifted by linear or angular displacements of the rotor axis in the same way as the signal pulses of the resonator pick-up. The distance between the two remains constant and is readily measured digitally.

The first experimental model using this optical pick-up had an operating threshold of 0.05°/h. This is the root mean square value of the fluctuations of successive zero-signal readings. The fluctuations recorded during several hours had an r.m.s. value of 0.1°/h. As no additional damping was provided, the time constant of the gyro was 23 sec. It should be mentioned that the rotor was rotating at 1500 rev/min only, in the bearings of a $15 motor, to the shaft of which it was mounted. All parts of the gyro were manufactured with low-precision tools. The resonator was suspended by one flat spring on each side of the resonator bar. As the respective planes of these springs were orthogonal, they combined high flexural stiffness with low torsional stiffness. The latter is required to match the natural frequency of the resonator to the low excitation frequency, which is equal to the frequency of rotation.

The comparatively long optical paths in the pick-up system, just described, limit the miniaturisation of the gyro. In order to reduce its size and at the same time improve the performance of the pick-up, we have to sample the inclination of the resonator not at two points of a revolution only, as before, but continuously, if possible, to increase the information rate. This consideration has led to a different optical concept.

There are no objections to a resonator in the shape of a cylindrical ring (Fig.4), which can oscillate about one of its diameters. Slots with an inclination of 45° are cut into the lateral area of the cylindrical ring. The light of a bulb, or other light source, in the centre of this ring passes through these slots onto a photo-transistor, which, due to the rotating of the ring, generates electrical pulses, the phase of which is modulated by the oscillating movement of the ring. To eliminate quasi-stationary displacements of the rotor axis, we rigidly attach a second ring to the rotor, monitored by a photo-transistor, and use the distance of the pulses from the two transistors for subsequent processing, which has to take the sign of the pulse distance into account, i.e. which pulse occurs first. There are no lenses or mirrors in the pick-up; all that is required is a light source, two photo-transistors and some electronics.

The pick-up worked well, but the gyro as a whole did not, and could not improve on 0.3°/h. To find out why our first gyro was nearly ten times better, we must look at the differential equation of the oscillations of the resonator. Time does not permit derivation of this equation here, but the assumptions made are that the inclination ϑ of the resonator is linearly, and only, dependent on the rate of rotation.

It is assumed that the resonator oscillates about a torsional axis passing through its centre of gravity; otherwise the gyro becomes sensitive to linear accelerations and, for a given eccentricity, this sensitivity is equal to that of conventional rate gyros. Further conditions are (i) the inclination ϑ must be small, i.e.,

$$\vartheta \ll 1 , \qquad (1)$$

(ii) the three components of the rate vector must be small and of second order as compared to the rotation ω, i.e.,

$$\Omega_1^2,\ \Omega_2^2,\ \Omega_3^2,\ |\Omega_1\Omega_2| \ll \omega^2 , \qquad (2)$$

and (iii) the component Ω_3 in the direction of the rotor axis must be small and of first order compared to the rotation ω and damping coefficient β, i.e.,

$$|\Omega_3| << \omega, \beta . \tag{3}$$

Then the differential equation is

$$\ddot{\vartheta} + 2\beta\dot{\vartheta} + \left[\omega_0^2 + \frac{C-B}{A}\omega^2\right]\vartheta = \left[\frac{A-B+C}{A}\omega\Omega_1 - \dot{\Omega}_2\right]\cos\omega t + \left[\frac{A-B+C}{A}\omega\Omega_2 + \dot{\Omega}_1\right]\sin\omega t. \tag{4}$$

A, B and C are the principal moments of inertia of the resonator and ω_0 is the natural frequency of the non-rotating resonator; ω_0 is determined by the spring restraint in the resonator suspension only. There are two periodic excitations with rotation frequency ω, and the resonator is tuned by making the coefficient of ϑ equal ω^2. It is not possible to make $C - B = A$, $(C-B)/A = 1$, to eliminate the need for ω_0 completely and thus provide automatic tuning for all rotational frequencies. However, for a resonator, which is flat in the direction of the rotor axis,

$$C - B \simeq A \tag{5}$$

and ω_0 is only small. To reduce the dependence on the rotational frequency, the resonator should be as flat as possible.

We may forget angular accelerations in (4) if

$$\left|\frac{\dot{\Omega}_1}{\Omega_2}\right|, \left|\frac{\dot{\Omega}_2}{\Omega_1}\right| << \omega . \tag{6}$$

Then, using the approximation (5), the differential Equation (4) reduces to

$$\ddot{\vartheta} + 2\beta\dot{\vartheta} + \omega^2\vartheta = 2\omega(\Omega_1\cos\omega t + \Omega_2\sin\omega t) , \tag{7}$$

with the stationary solution

$$\vartheta = \frac{1}{\beta}(\Omega_1\sin\omega t - \Omega_2\cos\omega t) . \tag{8}$$

There are two oscillations, with a mutual phase shift of 90°, each of which is determined by a different component of the angular velocity vector. As the two oscillations are easily separated by phase-sensitive detection, the Rotor-Vibragyro is a two-axis rate gyro.

The conditions specified do not differ from those commonly used in the theory of conventional rate gyros. For the latter we have to replace the rotational frequency ω in (2) by the natural frequency ω_G of rotor and gimbal frame about the output axis, and in (3) we have to replace β by the ratio ω_G^2/ω. Both β and ω_G^2/ω limit $|\Omega_3|$ and, as both are the reciprocal of the respective sensitivities (cf. Equation (8)), one cannot increase these limits without decreasing the sensitivities by the same amount in either case.

The conditions are equal for both types of rate gyros, but are they always satisfied? Operated under normal conditions, as a rate gyro with sufficient damping, a Rotor-Vibragyro, with resonator suspended in its centre of gravity, can only be disturbed by torsional oscillations of the rotor shaft about one of its diameters. These cause angular accelerations, which do not satisfy condition (6). In the presence of such vibrations we cannot neglect the acceleration terms and have to use Equation (4). For simplicity we consider inputs about one axis only.

Torsional oscillations are composed of periodic functions

$$\Omega_1(\tilde{\omega}) = \tilde{\Omega}(\tilde{\omega})\cos\tilde{\omega}t , \qquad \dot{\Omega}_1(\tilde{\omega}) = -\tilde{\omega}\tilde{\Omega}(\tilde{\omega})\sin\tilde{\omega}t , \tag{9}$$

Fourier components, which are to be considered as the input of the gyro. Because of the factors $\sin \omega t$ and $\cos \omega t$ they excite the resonator with frequencies $\tilde{\omega} + \omega$ and $\tilde{\omega} - \omega$. As we have tuned to ω, only vibrations with frequency $\tilde{\omega} = 2\omega$ become effective. By inserting (9) into (4), it can be seen that the effect of the velocity term is opposed to that of the acceleration term, reducing the unwanted output more and more as the principal moment of inertia A approaches $C - B$. We thus find a flat resonator to be desirable, not only to become more independent of rotational frequency but to reduce the effect of torsional vibrations as well.

This explains why the first Rotor-Vibragyro, with a resonator consisting essentially of a flat mirror, yielded a lower operating threshold than the second one, which used a cylinder extending 5 mm in the direction of the rotor axis. The sensitivity to torsional vibrations appears to be responsible for the operational threshold of both Rotor-Vibragyros.

Our current investigations are aimed not only at a reduction of these disturbances by flattening the resonator, but at their elimination by the use of two resonators, which are allowed to oscillate about two mutually perpendicular axes (Fig.5), both of which are perpendicular to the rotor axis. The two resonators are subjected to the same input rates, torsional vibrations as well as steady rates of turn. Calculations show that steady rates near zero frequency and rates oscillating with frequency 2ω are treated differently by the two resonators and a rough explanation of this is as follows. Torsional vibrations of frequency 2ω about a definite input axis cause the input rate to be a sinusoidal function of the rotor position with a fixed phase. This function has its maxima, for instance, at two fixed angular positions of the rotor, 180° apart from each other. In such a rotor position, the two resonator axes are at different angles to the input axis, thus receiving a different share of the input. On the other hand, the oscillations of the resonators due to steady rotations show a phase shift of 90° only, their amplitudes being equal. Calculations show that the amplitude of one of the resonators is proportional to the *sum* of steady and vibrational rates, while the amplitude of the other one is proportional to the *difference* of the two rates. This allows us to eliminate the torsional vibrations and retain the rates of rotation near zero frequency, which we intend to measure.

Experiments prove that the use of two resonators reduces the effect of rotational vibration of the rotor axis, and thus reduces the noise level of the Rotor-Vibragyro considerably. Reduction by a factor of 6 has been attained, with inadequate compensation, and this will be improved in the near future.

It may be concluded that the Rotor-Vibragyro is much less expensive than conventional gyros of comparable performance, because it does not require precision bearings for its rotating parts.

The concept of the Rotor-Vibragyro involves two points which are typical of this gyro:

(i) For transferring the signal from the rotor to the vehicle, two different optical pick-up systems have been described.

(ii) To eliminate the effect of torsional vibration of the rotor axis, causing a random drift of less than 0.3°/h with very poor bearings, the use of two crossed resonators has proved effective.

To improve mechanical and thermal stability it now is necessary to integrate the experience of the industry in the field of gyro technology into the development of the Rotor-Vibragyro.

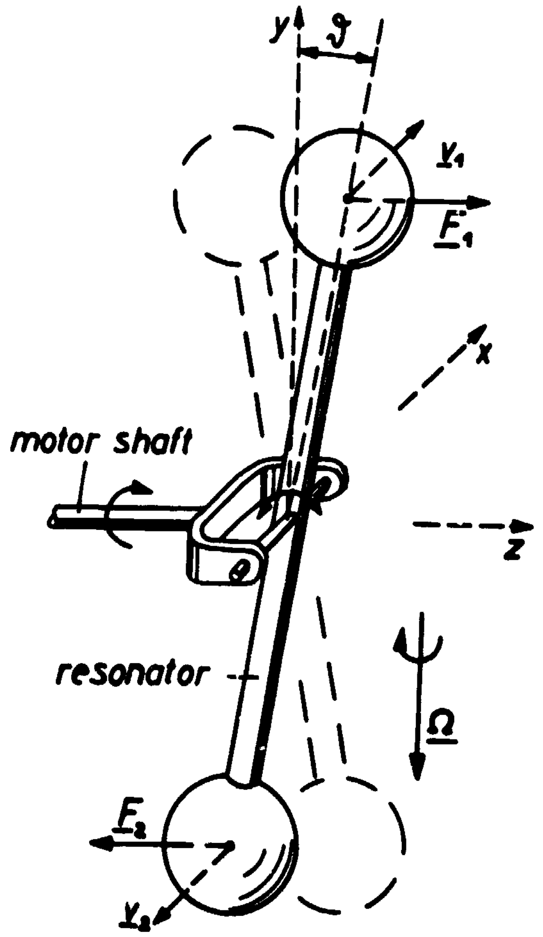
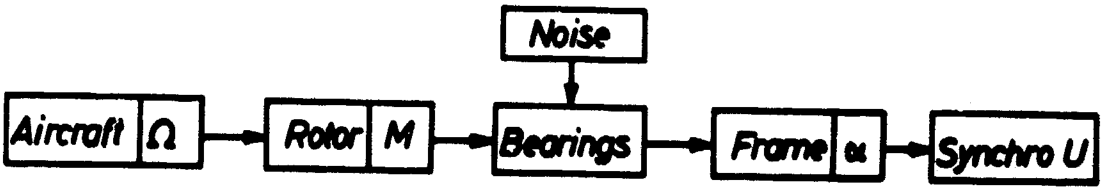
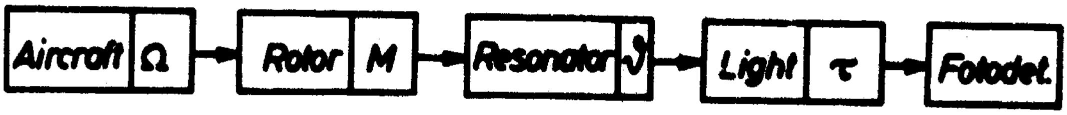


Figure 1



Conventional Rate Gyros



Rotor - Vibragyro

Figure 2

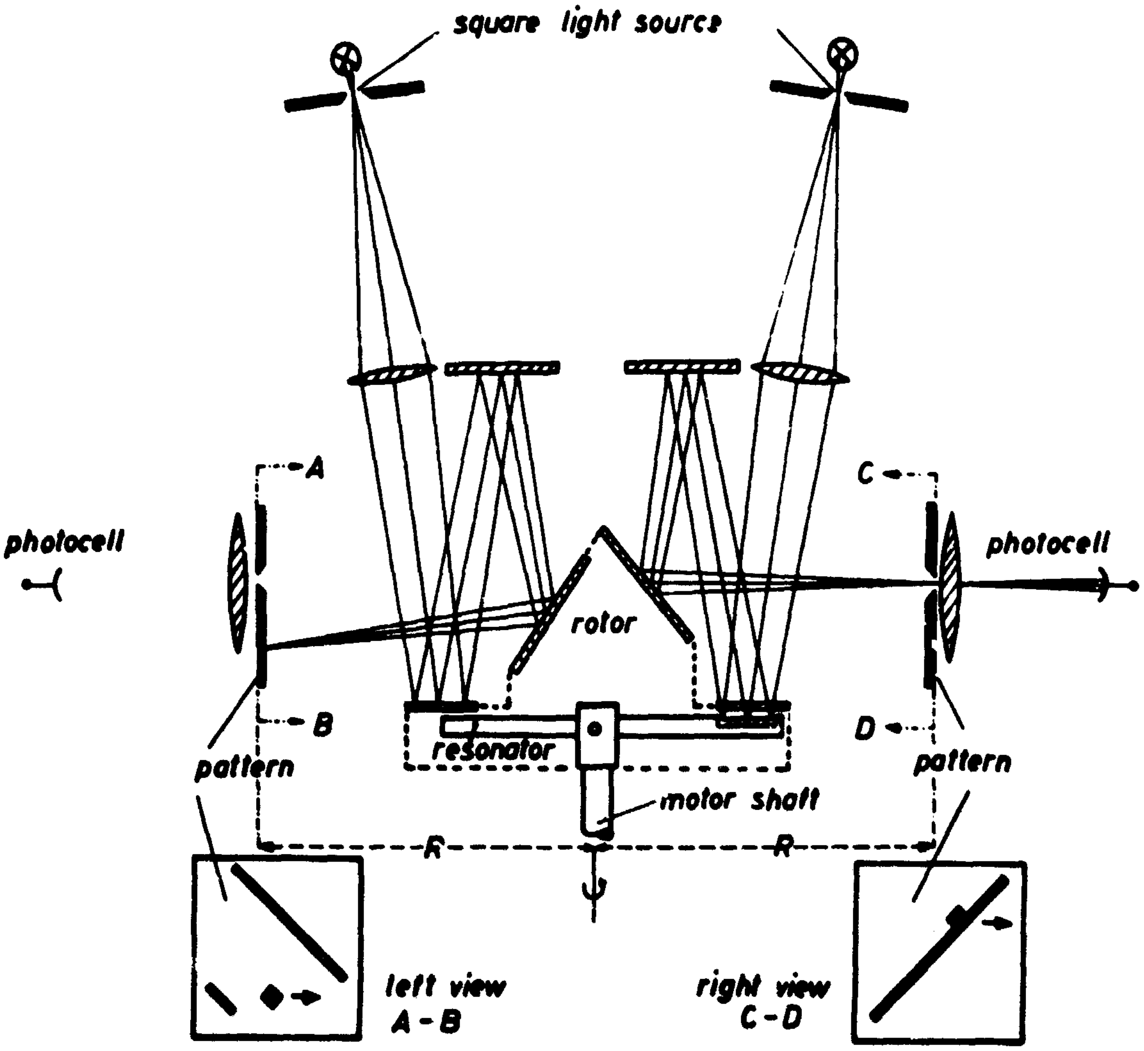


Figure 3

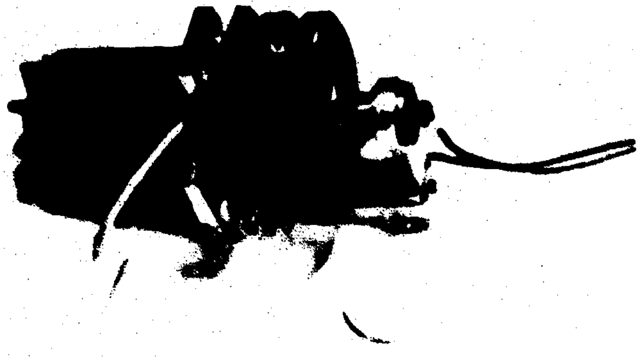


Figure 4

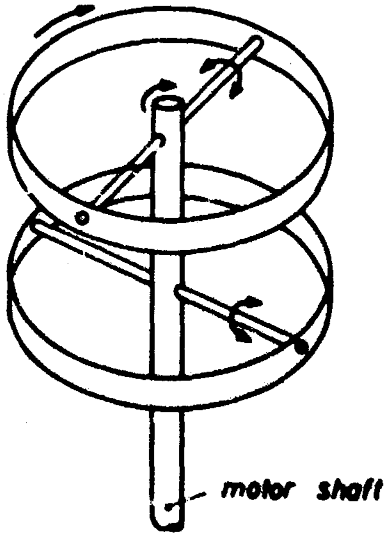


Figure 5

THE OSCILLOGYRO

by

R. Whalley and D.W. Alford

Ferranti Limited
Bracknell, Berkshire
England

THE OSCILLOGYRO

R. Whalley and D.W. Alford

1. INTRODUCTION

The paper describes a new type of gyroscope suitable for stabilising a platform in an inertial navigation system or for use as a heading reference. The main feature of the instrument is its inherent simplicity, which promises low cost manufacture coupled with high reliability. In its simplest form the gyro consists of a bar supported at its centre on a single-degree-of-freedom elastic hinge (Fig. 1). The bar and hinge assembly are fixed to a motor shaft which is rotated at an accurately controlled speed. A rigorous dynamic analysis of the instrument has been published previously (see Reference*1). In this paper the operation of the instrument is described, using simplified mechanical concepts.

2. DESCRIPTION OF OPERATION

The simplest way of visualising the behaviour of the bar is to consider it in terms of two well known dynamic effects; firstly, dynamic unbalance and, secondly, the gyroscopic torques.

It is convenient to consider the bar as a system of point masses joined with a weightless rod (Fig. 2). Consider the bar initially at some angle not perpendicular to the spin axis. As the driving shaft begins to rotate there will be a powerful torque caused by the centrifugal force acting on the masses. This will cause the masses on the bar to swing out until each mass lies in the same plane of rotation. It will be seen that in this position the centrifugal forces act along the axis of the bar and the bar will continue to rotate in one plane.

Consider now the bar spinning in this configuration. If the driving shaft is displaced through a small angle, gyroscopic forces will cause the bar to tend to continue to rotate in a fixed plane independently of the movements of the driving shaft. The elastic hinge will introduce a torque which will cause an unwanted precession. There is, however, a negative stiffness effect which arises from the dynamics of the system and which is proportional to the square of the rotation speed. The basic principle of the instrument is to arrange the rotation speed and the geometry of the bar so that the dynamic torque exactly cancels the spring torque. The behaviour of the bar will then approximate to that of a freely supported gyroscope with no restraints. The nature of this dynamic negative stiffness torque can be understood from Figure 3. It will be noted that the representation of the bar by point masses has been extended to include the mass distributed along the spin axis.

Let the driving shaft be tilted through a small angle α. This will introduce a cyclic component of rotation about the X-axis of the bar with maximum value

$$W_x = \Omega \sin \alpha .$$

Velocity of masses distributed along Z-axis.

$$V \simeq Z\Omega\alpha .$$

Coriolis acceleration,

$$A = 2\Omega V = 2\Omega^2 Z\alpha .$$

Dynamic torque about Y-axis,

$$q_D = 2M(2\Omega^2 Z^2\alpha) = 4MZ^2\Omega^2\alpha .$$

Spring torque about Y-axis due to elastic suspension,

$$q_S = -K\alpha .$$

Summing torques,

$$K\alpha = 4MZ^2\Omega^2\alpha ,$$

which gives the tuned condition when

$$\Omega^2 = \frac{K}{4MZ^2} .$$

The plane in which the bar is rotating can be precessed by applying a torque about a perpendicular axis, as with a normal gyroscope. In practice an impulse of torque is applied to the bar a quarter of a revolution before the axis of precession is reached.

We have discussed the torque arising from the elastic hinge when the driving shaft is misaligned and the dynamic torque which compensates this. There is, however, another torque which arises from the damping forces. When the bar is rotating in a fixed plane and the driving shaft is misaligned by a small angle (α), the bar will appear to oscillate on its hinge $\pm\alpha$ in one revolution relative to the fork and the other rotating parts of the instrument. This motion introduces a damping torque mainly due to windage which will precess the gyro about the axis of misalignment and in a direction to reduce the misalignment to zero. It will be noted that a similar effect appears in all free rotor gyros when the driving shaft is misaligned. The gyro is precessed by the damping forces at a rate proportional to the misalignment and the misalignment is reduced exponentially with time. The damping torque represents an unwanted disturbance of the gyro and, in the case of the Oscillogyro, it is reduced to a very small level by mounting the bar in a rotating container which is sealed and evacuated.

It is important to realise that, in adopting a single-degree-of-freedom suspension, the ability of the gyro to control a platform about two axes has not been lost. Pick-offs can be arranged in two pairs to measure the proximity of the bar to a reference plane on the instrument in each of four quadrants. In this way one pair of pick-offs will measure displacement of the gyro about one axis and, a quarter of a revolution later, the other pair will measure the displacement about an axis at right angles. The instrument can be regarded as a two-degree-of-freedom gyro working on a time-sharing basis.

The advantages of the Oscillogyro can now be considered. The sensitive mass is inherently simple and this minimises instability of the mass unbalance torques. There is, for example, an obvious advantage, compared with more conventional instruments, in having the rotor bearings (with their possibilities of mass shifts due to wear) outside the sensitive mass. When the gyro is used in the spin axis vertical configuration, it will be seen that there can be no drift due to unbalance since the unbalance vector is rotating with the rotor, and drift (relative to a fixed frame of reference) is commutated to zero. It will be seen, similarly, that a bias torque from the suspension system cannot cause drift about fixed axes. From the cost point of view the main advantage is the inherent simplicity of the instrument; the elastic hinge is simple and cheap to manufacture and there is no requirement for extreme cleanliness. The dynamic compensation for spring stiffness allows the use of a very robust elastic suspension without introducing excessive demands on the servo follow-up system.

The sensitive element has been described in terms of a bar and it is important to look briefly at the reasons for this, i.e., the factors which control the shape of the sensitive mass. When the gyro is running at its correct speed the torque arising from the spring stiffness due to a small misalignment (such as an error in the platform servo) will be exactly compensated for by the dynamic torque. However, if there is an error in speed (or in the stiffness of the hinge) there will be an unwanted torque which will cause drift. It will be seen that there is an advantage in minimising this by keeping the uncompensated suspension stiffness as low as possible consistent with adequate shock capability of the suspension. The tuning equation

$$\Omega^2 = \frac{K}{4MZ^2}$$

shows that, for K to be kept small, either the speed or the axially distributed mass MZ^2 must be minimised. There are good reasons for running any design of gyroscope at as high a spin speed as possible (for example, to maintain a good ratio of angular momentum to mass), and with the Oscillogyro there is the added requirement that the rotation speed must be kept sufficiently high to provide a good servo performance. Now this means that the MZ^2 term should be minimised and in practical designs the sensitive mass tends to have the characteristic shape of a bar with a large ratio of diameter to thickness.

At this point it is perhaps appropriate to consider the similarity of the Oscillogyro with the tuned Hooke's joint elastic hinge gyro. The gimbal of the Hooke's joint gyro can be regarded in the same way as the bar of the Oscillogyro in tuning out the spring torques. This has been emphasised in Figure 4, where a rotor has been added around the bar, which is partly decoupled from the bar by a second elastic hinge system.

At the risk of over-simplification it can be said that the angular momentum has been increased by the addition of the ring without significantly increasing the dynamic torque - since the ring will have virtually no angular velocity about the X-axis. The tuned speed will therefore depend largely, as before, on the dimensions of the bar and the stiffness of the hinge. However, for a given speed and servo error, there will be a lower drift rate, simply because the angular momentum has been increased.

We can see in a general way that the Hooke's joint suspension will allow gyroscopes to be designed with a lower drift rate for a given speed and servo error, but with this there will be a reduction in shock loading capability.

In our experience the requirement for a good servo performance is set as much by the damping torques as by the errors in tuning. The advantage of simplicity in the use of a single bearing, we believe, may outweigh the theoretical advantages of the two-degree-of-freedom approach. It may well be that the features of the two-degree-of-freedom Hooke's joint design can best be exploited in low acceleration applications.

3. THE PRACTICAL INSTRUMENT

Figure 5 shows the main functional parts of one form of Oscillogyro. The instrument has been designed with ease of servicing in mind and it can be disassembled into three main sub-assemblies, the drive motor, the sealed oscillobar container and the resolver and pre-amplifier assembly. The drive motor is a hysteresis type and is driven from a constant frequency supply to provide a constant speed drive. This is necessary to maintain the tuned condition and also to provide sufficient accuracy of precession by the torque motor. The oscillobar is mounted on a cross-leaf elastic suspension. It is balanced and tuned and the container is then sealed and evacuated.

The pick-offs and torque motors are both electrostatic devices. The pick-offs form a capacitance bridge circuit operating at 200 kHz. Insulated plates are mounted on the rotating container so that there is a small gap between the plates and the oscillobar. Four plates are used, one pair at each end of the bar. These plates are connected to insulated forks outside the sealed container which form the rotating member of a capacity resolver. The stationary member of the resolver is divided into segments which are connected to pick-off and torque motor circuits. As the container rotates, the insulated plates each side of the bar are capacitatively coupled in turn to the torque motor circuit and to the pick-off circuit, first for one channel and then for the other.

The capacitative pick-off has a high source impedance and a pre-amplifier is necessary close to the pick-off. In this design the pre-amplifier is constructed using field effect transistors and it is mounted on the resolver stator.

The rotating plates are coupled to the pick-off circuit twice per revolution. The output of the pick-off consists, therefore, of a train of pulses which are, in effect, envelopes of a.c. at the pick-off excitation frequency, the amplitudes of which represent the input angle, i.e., the angular movement of the gyro case. The two axes can be brought out on one line through one head amplifier or, alternatively, on separate lines (see Figure 6).

These pulses have to be rectified in a phase-sensitive rectifier and then passed through a sample-and-hold circuit before being used to operate the platform servos. The function of the sample-and-hold circuit is to average the level of the pulse and to hold it as an error input to the servo until the next pulse presents itself. The time between successive pulses represents a lag in the servo loop, but this is less than 4 ms for rotation speeds in excess of 8000 rev/min and does not create a serious problem in the design of the platform servos.

4. CONCLUSIONS

The Oscillogyro belongs to the new family of tuned elastically supported gyroscopes, several examples of which are currently under development at different research centres. The approach offers an inherent immunity from a number of the error sources which affect more conventional gyroscopes, together with a basic simplicity of construction. It promises low cost manufacture coupled with high reliability.

ACKNOWLEDGEMENTS

The work with which the authors have been associated derived from an idea proposed by Mr J.St Leger Philpott and Mr J.H.Mitchell, who were at the Royal Aircraft Establishment, Farnborough during the 1939-45 war (Ref.2).

The initial experimental and design work was carried out by Mr J.W.Barnes, first at RAE and later at Ferranti Limited.

The authors acknowledge with thanks the permission of the Ministry of Technology and the Directors of Ferranti Limited to publish this paper.

REFERENCES

1. Whalley, R., et al.	*The Oscillogyro.* Journal of Mechanical Engineering Science, Vol.9, No.1, February 1967.
2. -	British Patent No.599826.

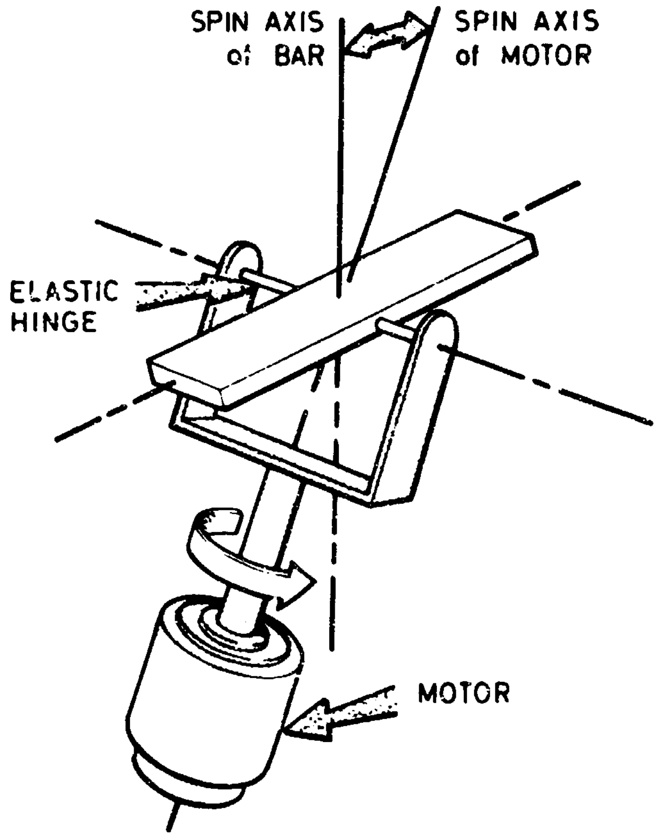


Fig. 1 Basic Oscillogyro

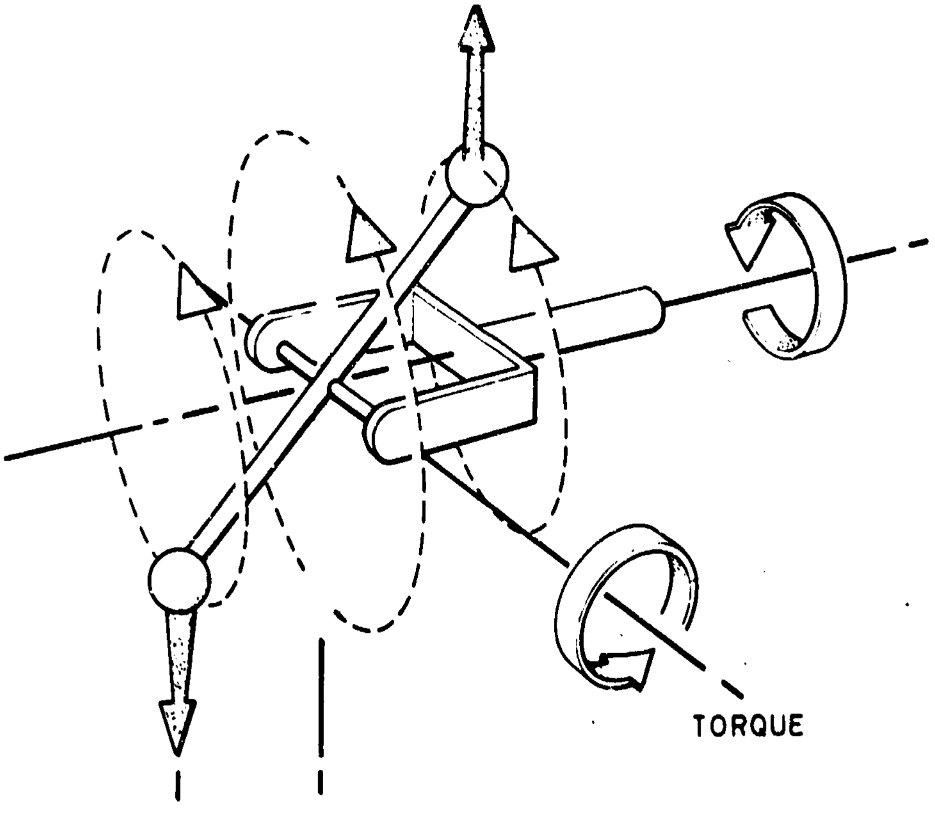


Fig. 2 Illustrating motion of bar

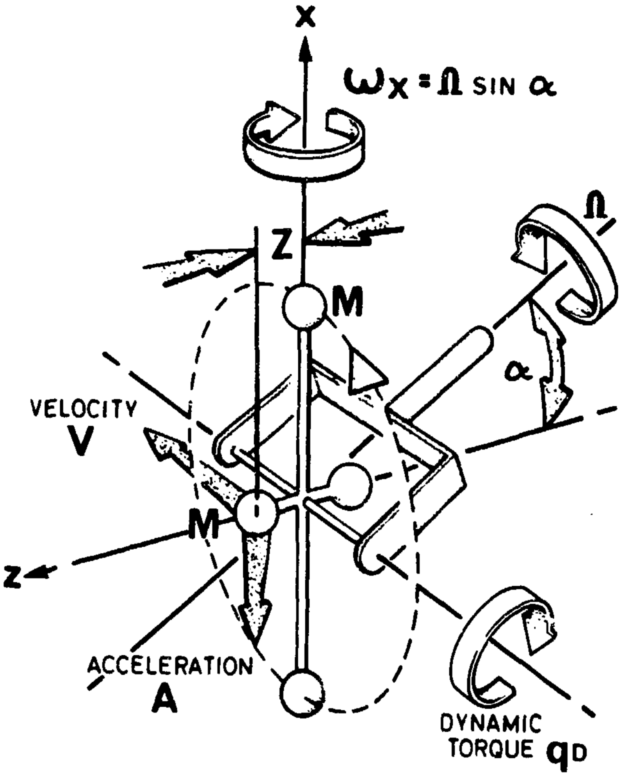


Fig. 3 Principles of dynamic tuning

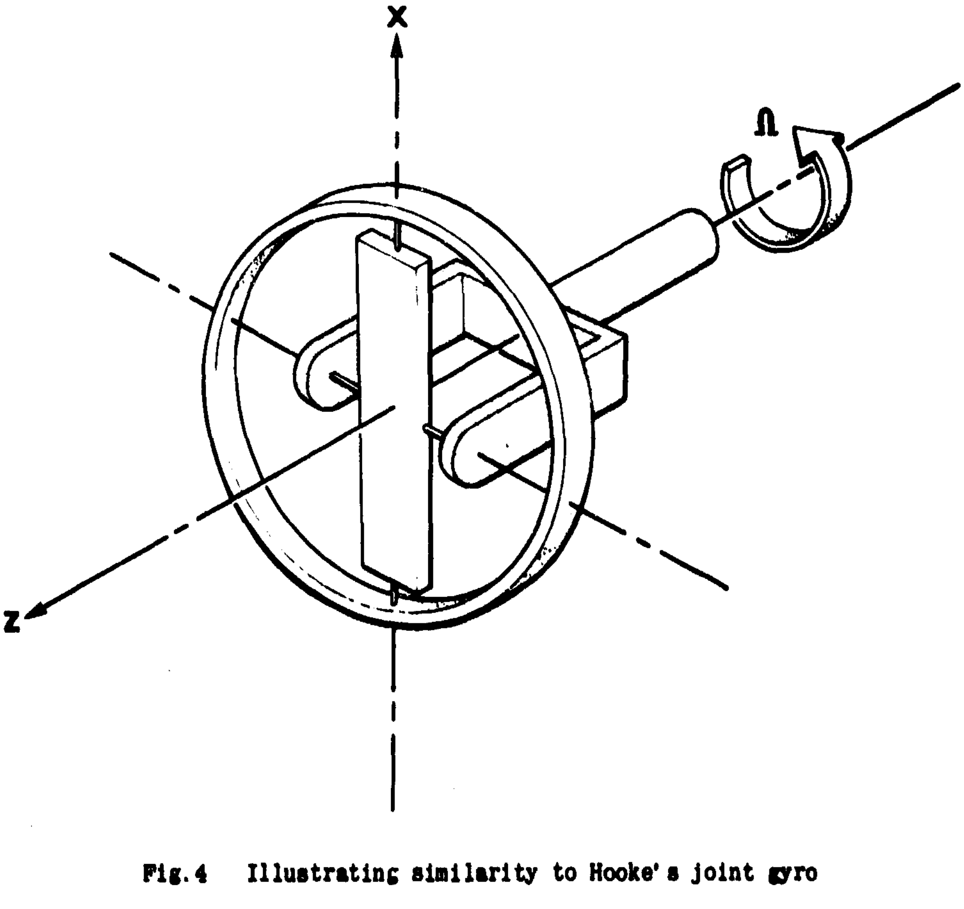


Fig. 4 Illustrating similarity to Hooke's joint gyro

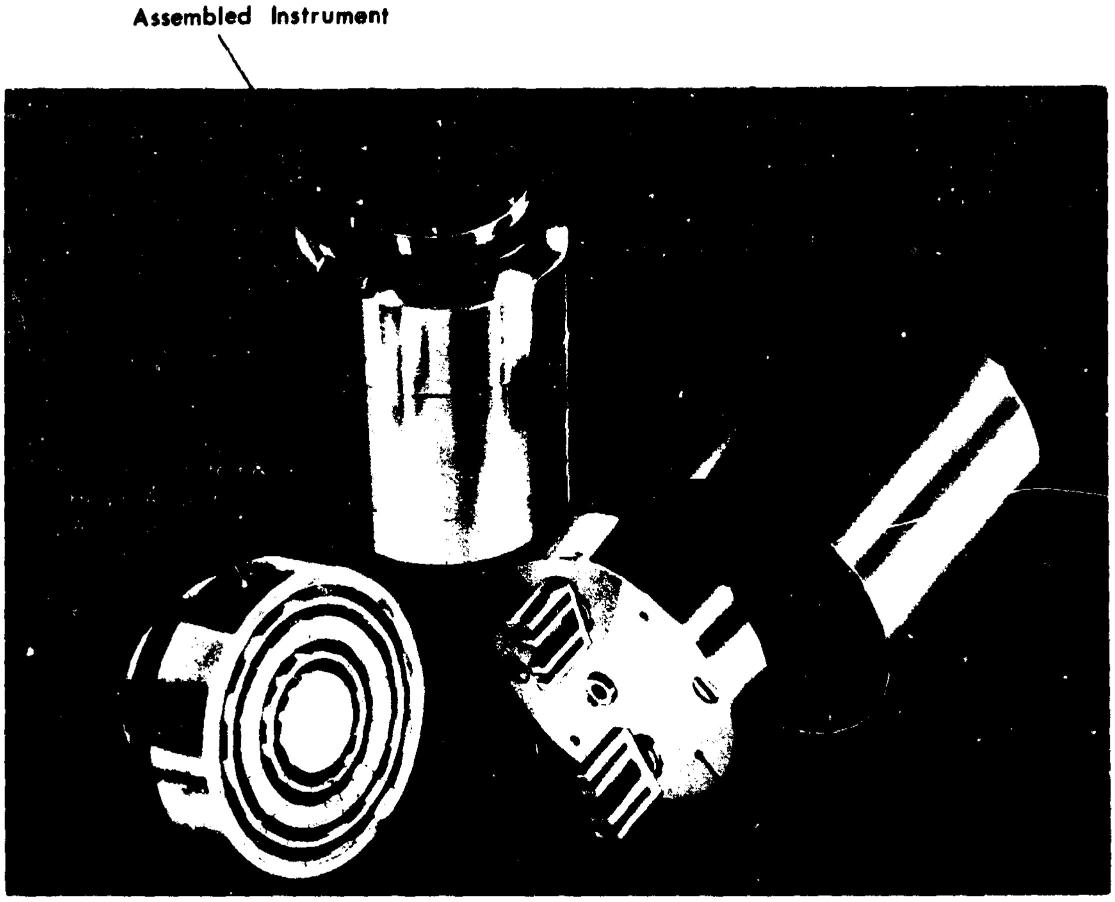


Fig. 5 Sub-assembly breakdown of experimental prototype

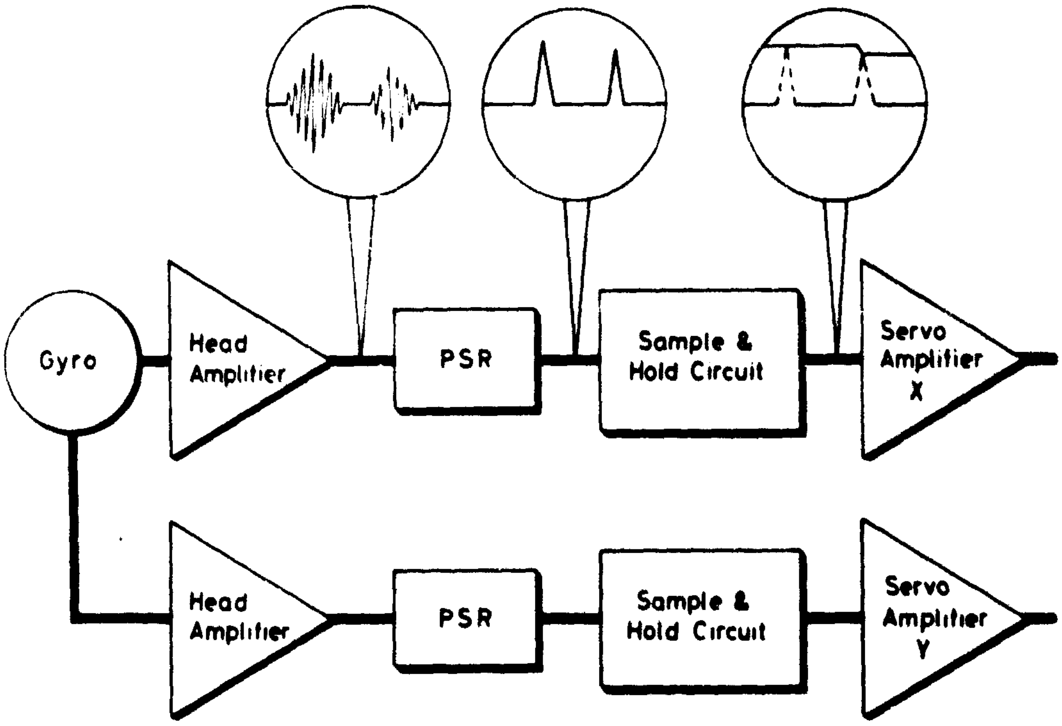


Fig. 6 Schematic of pick-off outputs

THE PIPA

(PULSED INTEGRATING PENDULOUS ACCELEROMETER)

by

George J. Bukow

Massachusetts Institute of Technology,
Cambridge, Massachusetts, USA

SUMMARY

The PIPA (Pulsed Integrating Pendulous Accelerometer) used in The Apollo Inertial Measurement Unit is a moderate cost, high performance single-degree-of-freedom, specific force integrating receiver, operating closed loop as a torque restrained pendulum. The instrument is one of a class of floated inertial components, consisting of a pendulous float magnetically suspended in a viscous fluid, signal generator and torque generator microsyns, and associated electronics and calibration modules.

The paper examines in detail the theory, fabrication, and operation of the instrument. The float dynamics of a binary (two-state) torquing loop are derived and the results of a digital simulation relating the design parameters to the oscillatory mode and response of the instrument are presented.

The stability of the electromagnetic torquing circuit is shown to be one of the primary limitations of the present instrument design, and the factors that affect this stability are examined. Other areas discussed include: torque errors resulting from fluid impurities, floatation problems, failure modes, and power requirements.

The paper concludes with a summary of present performance capabilities and recommendations for improvement. Suggested changes would substantially improve the present performance and make the instrument suitable for longer and more demanding missions.

THE PIPA
(PULSED INTEGRATING PENDULOUS ACCELEROMETER)

George J. Bukow

1. INTRODUCTION

The PIPA (Pulsed Integrating Pendulous Accelerometer) used in the Apollo Inertial Measurement Unit is a single-degree-of-freedom specific force integrating receiver, operating closed loop as a torque-restrained pendulum. The instrument consists of a pendulous float magnetically suspended in a viscous fluid, signal generator and torque generator microsyns, and associated electronics and calibration modules. A schematic diagram of the Pulsed Integrating Pendulum (PIP) is shown in Figure 1.

The pendulous float is a hollow beryllium cylinder with built-in mass unbalance. Ferrite rotors are mounted at each end of the float, supporting the magnetic suspension and the microsyn functions. The float is surrounded by heavy fluid which provides both a buoyant support and viscous damping.

An applied specific force along the input axis of the pendulum causes a rotation about the output axis. This rotation is sensed by the signal generator and is used as an error signal to initiate a response from the torque generator.

A cutaway view of the complete pendulum is shown in Figure 2. The pendulous mass is shown mounted on the periphery of the float and extends into a small groove in the damping block. The width of the groove allows maximum rotation of the float of $\pm 1^\circ$ of arc. The magnetic suspension stators are positioned opposite the inside portions of the rotors, forming part of a coaxial structure with the torque generator on one end, and with the signal generator on the opposite end of the instrument. Four bellows provide volumetric compensation of the damping fluid between 10° and $71^\circ C$. On the end of the instrument, two aligning rings permit accurate positioning on the PIPA's input axis relative to the mounting surface. The instrument is aligned to the mounting ring before incorporation in the inertial measuring unit (IMU).

The overall block diagram, showing the pendulum within its control loop, is shown in Figure 3. The signal from the signal generator is amplified and used as an input to a sampler (called the interrogator) which samples the signal at discrete times determined by the guidance computer. If the signal exceeds either the positive or negative threshold of the interrogator, a command pulse is sent to the torque generator switch. The switch, in turn, commands the proper polarity current to the torque generator of the instrument, nulling out the error signal from the signal generator.

The switching times are precisely controlled by the switching pulse train, and the torquer current magnitude is controlled by an ultrastable d.c. current loop. Torque is applied to the float in discrete increments, and for each increment a pulse is generated and sent to the computer, providing a continuous monitoring of the total torque applied to the instrument.

2. BASIC PARAMETERS

2.1 Scale Factor

The differential equation for the general response of the pendulous float to input accelerations and restraining torques from the torque generator may be written as

$$J\ddot{A}_f(t) + C\dot{A}_f(t) + KA_f(t) = M_{TG}(t) - ml\, a_{in}(t) \,. \tag{1}$$

where

$A_f(t)$ = angular position of the float relative to its zero or reference position (radians)

J = float angular moment of inertia (gm cm^2)

C = viscous damping (dyne cm sec/rad)

K = elastic restraint (dyne cm sec/rad)

$M_{TG}(t)$ = torque applied by the torque generator (dynes)

ml = pendulosity (gm cm)

$a_{in}(t)$ = acceleration input (cm/sec^2) .

The elastic restraint term is due to the asymmetries in the suspension and microsyn circuits and is small compared to the magnitude of the applied torque. Thus it may be neglected for most considerations. One exception is the calculation to determine the minimum threshold acceleration of the instrument, covered in Section 2.3.

Integrating Equation (1),

$$J\dot{A}_f(t) + CA_f(t) = \int_0^t M_{TG}(t)\,dt - ml\int_0^t a_{in}(t)\,dt + C \,. \tag{2}$$

The constant term results from the initial conditions and is equal to $J\dot{A}_f(0) + CA_f(0)$. $M_{TG}(t)$ is applied incrementally in precisely controlled bits in either the positive or negative direction. For the present, these increments of torque will be assumed equal, in both the positive and negative directions, and of magnitude $M_{TG}\,\Delta T$; here M_{TG} is the average magnitude of the applied torque and ΔT is the length of time each increment of torque is applied. Thus the applied torque integral may be written

$$\int_0^t M_{TG}(t)\,dt = \sum_{t=0}^{t=t} \pm (M_{TG}\Delta T) \,.$$

Incorporating these changes into Equation (2),

$$J(\dot{A}_f(t) - \dot{A}_f(0)) + C(A_f(t) - A_f(0)) = \sum_{t=0}^{t=t} \pm (M_{TG}\Delta T) - ml\int_0^t a_{in}(t)\,dt \,. \tag{3}$$

The left-hand side of Equation (3) represents a change in momentum stored within the float. Under controlled conditions this momentum change may be assumed equal to zero, further simplifying the equation to

$$\sum_{t=0}^{t=t} \pm (M_{TG}\,\Delta T) = ml\int_0^t a_{in}(t)\,dt \,. \tag{4}$$

Equation (4) presents the basic relation between the input acceleration and the applied restraining torque. The integrated acceleration, or velocity increment, obtained over any given period of time, is equal to $n(M_{TG}\,\Delta T/ml)$, where n is the net number of torque pulses occuring over that interval of time. The term $(M_{TG}\,\Delta T/ml)$ is the velocity equivalent of one torque pulse and is defined as the scale factor of the instrument, having the units of cm/(sec pulse).

2.2 Binary Operation

The torquing loop employed with the PIPA is a two-state (or binary) loop, with torque applied continuously in either positive or negative directions. Therefore, even with no input acceleration, current is switched alternately from positive to negative torque windings and the float is maintained in a continual state of oscillation. The parameters of the instrument determine which state of oscillation, or mode, the float will maintain. If "m" pulses occur alternately in each direction, an m:m mode is defined to exist. A complete derivation and discussion of the dynamic response of the instrument is presented in Section 4.

Under a no-input condition, the equation of equilibrium becomes

$$\sum_{t=0}^{t=t} (M_{TG}^{+}\,\Delta T - M_{TG}^{-}\,\Delta T) = 0\,, \tag{5}$$

assuming no change in float storage. In general, $M_{TG}^{+} \neq M_{TG}^{-}$, and the number of positive pulses will not equal the number of negative pulses over any given period of time.

If n_1 is the number of positive torque pulses applied during time t and n_2 is the number of negative pulses, the following relations may be obtained:

$$\frac{n_1}{n_2} M_{TG}^{+} = M_{TG}^{-} \tag{6}$$

$$\frac{M_{TG}^{+} + M_{TG}^{-}}{2} = M_{TG}^{+}\,\frac{n_2 + n_1}{2n_2} \tag{7}$$

$$\frac{M_{TG}^{+} - M_{TG}^{-}}{2} = M_{TG}^{+}\,\frac{n_2 - n_1}{2n_2}\,. \tag{8}$$

Dividing Equation (7) by Equation (8) and cross-multiplying yields

$$(n_2 - n_1)(M_{TG}^{+} + M_{TG}^{-}) - (n_2 + n_1)(M_{TG}^{+} - M_{TG}^{-}) = 0\,. \tag{9}$$

Multiplying by $\Delta T/2ml$,

$$(n_2 - n_1)\,\frac{M_{TG}^{+} + M_{TG}^{-}}{2ml}\,\Delta T - (n_2 + n_1)\,\frac{M_{TG}^{+} - M_{TG}^{-}}{2ml}\,\Delta T = 0\,. \tag{10}$$

Equation (10) presents the equilibrium relationship between the scale factor and the bias of the unit under a no-input condition. The term $(M_{TG}^{+} - M_{TG}^{-})\Delta T/2ml$ is similar to the scale factor defined in Section 2.1 and is termed the average scale factor. The differential term $(M_{TG}^{+} - M_{TG}^{-})\,\Delta T/2ml$ represents the imbalance existing between the positive and negative torquers and is defined as the bias.

2.3 Minimum Threshold

The minimum threshold of the instrument may be determined by again referring to Equation (1). For very small inputs the motion of the float may be treated as the sum of two terms: an oscillator response due to the alternate torquing of the positive and negative torquers, and a slow rotation of the float caused by the input acceleration. If the float moves far enough from null, one torquer remains on for one pulse length longer than normal to return the float to null.

In steady state, the float motion resulting from one pulse of torque will be equal to $(M/C)\Delta T$, since all of the energy of the pulse is used to move the float through fluid. Thus, the initial excursion of the float required to cause a torquer imbalance of one pulse is again $(M/C)\Delta T$. The torque produced by elastic restraint t_K for an excursion of $(M/C)\Delta T$ is

$$t_K = \frac{KM}{C}\Delta T \, . \tag{11}$$

The input acceleration must cause a torque on the float at least equal to t_K in order to be sensed by the instrument. Thus

$$ml a_{in} \geq \frac{KM}{C}\Delta T ,$$

$$(a_{in})_{min} = \frac{KM\Delta T}{Cml} . \tag{12}$$

3. DESCRIPTION OF OPERATION

3.1 Magnetic Suspension

The PIPA makes use of two quasi-elastic magnetic supports to properly position the neutrally buoyant float within the case. A cross-section of the magnetic structure used is shown in Figure 4. This figure depicts the unique feature of separated suspension and transducer functions using a single coaxial structure. Radial and axial centering is produced by action of the suspension magnetic field on the inner surface of the tapered rotor.

A schematic of the suspension circuit is shown in Figure 5. The four coils represent the effective inductances along the plus-and-minus X and Y axes. The capacitors are mounted in a potted module external to the PIP and provide the proper tuning to effect stable magnetic restoring forces on the rotor.

The underlying basis for the suspension circuit used may be simply stated as

$$F_R = \frac{\partial W_g}{\partial g} . \tag{13}$$

where

F_R = restraining force on the rotor due to a stator pole

W_g = magnetic energy in the air gap

g = air-gap length .

The air-gap energy is

$$W_g = \tfrac{1}{2}(\mathrm{MMF})_g^2 P_g , \tag{14}$$

where

$(\mathrm{MMF})_g$ = air-gap magnetomotive force

P_g = air-gap permeance .

Therefore

$$F_R = \tfrac{1}{2}(\mathrm{MMF})_g^2 \frac{\partial P_g}{\partial g} . \tag{15}$$

To establish a relationship between the suspension stiffness, or dF_R/dg , and the circuit parameter requires considerable algebraic manipulation[1]. The results of the derivation are summarized below. Defining the terms

$$Q_c = \frac{\omega_0 L_0}{R} , \tag{16}$$

$$Q_t = \frac{\omega_0 L + \omega_0 L_0 - 1/\omega_0 C}{R} . \tag{17}$$

and

$$f(Q) = \frac{1 + Q_t^2 - Q_c Q_t}{(1 + Q_t^2)^2} . \tag{18}$$

where

Q_c = the Q of the suspension coil

Q_t = the overall Q of the circuit

L_0 = self-inductance of the coil

L = leakage inductance

C = series capacitance

R = effective resistance

and ω_0 = excitation frequency .

$$\frac{dF}{dg} = \frac{L_0}{(g_0)^2} \frac{V^2}{R^2} f(Q) . \tag{19}$$

where

V = terminal voltage

g_0 - centered air-gap value .

In order to have a stable restoring force, $f(Q)$ must be negative and Q_t must be greater than zero. A family of curves showing the values of $f(Q)$ as a function of Q_t , for different values of Q_c , is shown in Figure 6. The maximum suspension stiffness occurs near the three-quarter-power point, where $Q_t = 1/\sqrt{3}$.

The normal operating point for the PIPA is the second (inductive) half-power point, shown on the curve. This point was chosen to insure a stable suspension for all possible values of g_0 resulting from axial motion of the float. When the float is not suspended it can move to an axial extreme. If the suspension is applied at such time, the value of g at one end of the instrument will be higher than normal, resulting in a smaller value of Q_1. This value of Q_1 must be kept larger than the minimum value required for adequate stiffness.

3.2 Signal Generator

Float rotations are sensed by an angle-to-voltage transducer called a signal generator. A cross-section view of the signal generator stator and rotor is shown in Figure 7. The primary coil of the signal generator (SG) is wound in series around each of the eight stator poles. The flux paths set up by these poles are typified by the flux paths shown in the figure. The SG secondary coils are wound such that the even-numbered poles are in phase with the primary, and the odd-numbered poles are 180° out-of-phase. If the contoured outside diameter of the rotor is perfectly centered and at zero rotation relative to the stator (as shown in the figure), the total signal induced in the secondary of the SG is zero. If a rotor (float) rotation occurs in either the positive or negative direction, a corresponding in-phase or out-of-phase output is obtained across the SG secondary, since the reluctance seen by the positive secondary poles is no longer equal to that seen by the negative poles.

The current in the SG primary may be written as $i_p = I_p \cos \omega_0 t$. The corresponding voltage induced in the secondary is

$$V_S(t) = K_S I_p \frac{d}{dt} [A_f(t) \cos \omega_0 t] , \qquad (20)$$

where K_S is a proportionality constant resulting from the circuit geometry. Expanding Equation (20),

$$V_S(t) = K_S I_p (\cos \omega_0 t) \dot{A}_f(t) - K_S I_p A_f(t) \omega_0 \sin \omega_0 t . \qquad (21)$$

The induced secondary voltage is filtered by a tuned second-order filter, as shown in Figure 8(a). The output of the filter lags the induced voltage by 90°. If the time of interrogation is at $t = 0$, the filtered rate term, $K_S' I_p \cos (\omega_0 t - 90^\circ) \dot{A}_f(t)$, will be crossing through zero and the total instantaneous output will be a function of only the float position. The phase relationships of the voltages and currents in the SG circuit are shown in Figure 8(b).

3.3 Torque Generator

Except for the way the coils are wound, the torque generator (TG) cross-section is identical to that of the SG. The position of the TG coils and resultant flux paths for positive torque are shown in Figure 9. Since only one torquer is on at any time, there are two minimum-energy rotor positions, one for positive torque and one for negative torque. The rotor always tends to line up at the minimum-energy position opposite the working poles. However, the frequency of switching allows only a minute oscillation about a center point midway between the minimum energy points. The torque produced is proportional to the square of the flux linking the rotor poles and thus is a function of the square of the torquer current.

3.4 Electronics

A portion of the overall control-loop block diagram detailing the electronics is presented in Figure 10. The filtered SG signal is amplified by the preamp and a.c. differential amplifier and sent to the interrogator. The interrogator utilizes two Schmidt triggers, synchronized by the interrogator pulse train from the computer, or system clock, to determine the polarity of the signal from the a.c. amplifier. The outputs from the Schmidt triggers control the state of two multivibrators, commanding either positive or negative torque from the switch.

The binary current switch controls the direction of the current flow from the current source, in accordance with the commands received from the interrogator. A switch pulse train synchronizes the switching of the second multivibrator stage in the interrogator such that the pulses of current sent through the torque windings are of equal width.

The current control loop is expanded in Figure 11. The output of the switch is sent to the calibration module. This module provides the necessary passive components to make the torquer coils of the PIP appear purely resistive, thus controlling switching transients and optimizing current stability. Bias adjustment is obtained by use of the resistive shunting circuit; and fine padding of the resistive loads R_1, in series with the torque windings, allows equalization of the time constants τ_1 and τ_2.

The voltage drop across the scale factor (SF) resistor is compared with a precision voltage reference (PVR). The error voltage is amplified by a high-gain d.c. differential amplifier and used to control the current magnitude in the loop.

3.5 Parameter Summary

Figure 12 summarizes the typical operating parameters for the Apollo PIPA. Values for torque constant, torque current, and scale factor are given for the two PIPA configurations used, Command Module (CM) and Lunar Module (LM). Different calibration modules are used to provide the difference in torque current between the two configurations.

The open-loop power includes that used for the suspension and SG circuits. The closed-loop power is the sum of the open-loop power and the power dissipated in the torque windings.

4. DYNAMIC RESPONSE

It has been shown[2] that pulse rebalance loops with nonlinear samplers can be treated with describing function techniques to predict the modes of oscillation that will occur for given operating parameters and input conditions. While such methods are helpful as general design tools, they do not readily allow adequate definition of the control loop to permit accurate analysis of an operating instrument. To obtain the required accuracy and definition, a piecewise linear approach was used in the analysis of the PIPA.

Between any two switching times, and with the assumption of some constant input acceleration, the PIPA float motion is easily described by a second-order differential equation. Representation of other portions of the control loop by linear transfer functions allows the voltage input to the interrogator to be calculated at each sample time. If the voltage seen by the interrogator exceeds the threshold voltage, a switching condition occurs and the torque reverses direction. The initial conditions may then be updated and the process repeated. In this manner the output pulse train of the instrument may be obtained, as well as its response to a particular set of initial conditions.

For the purpose of this derivation, the buoyant support, augmented by the magnetic suspension, is assumed to be infinitely stiff. Small deviations from this condition in an actual instrument will not materially affect the conclusions reached. The mathematical expressions for the different portions of the loop are presented in the following sections. The justification for the mathematical model derives from the discussion of the PIPA operation presented in Sections 1-3.

4.1 Equation of Motion

The basic equation of motion of the float under the assumption of zero input acceleration is

$$J\ddot{A}_f(t) + C\dot{A}_f(t) = M(t) . \tag{22}$$

The expression for $M(t)$ is

$$M(t) = M(1 - e^{-t/\tau})^2 \Big|_0^{n\Delta T} + M(e^{-t/\tau})^2 \Big|_{n\Delta T}^{\infty} , \tag{23}$$

where

τ is the time constant of the torquer current

M is the maximum torque applied

n is the number of torque pulses applied between switching times

ΔT is the length of one pulse of torque.

For normal operation at the instant of switching, the current in one torquer begins to decay to zero while that in the other begins its rise to maximum current. Thus, the total torque applied from time zero (the switching time) until the next instant of switching is

$$[M(t)]_{total} = M(1 - e^{-t/\tau})^2 - (e^{-t/\tau})^2 , \tag{24}$$

where the rise times in the two current loops are assumed equal and $M^+ = M^-$. Simplifying,

$$M(t) = M(1 - 2e^{-t/\tau}) . \tag{25}$$

Substituting in Equation (22),

$$J\ddot{A}_f(t) + C\dot{A}_f(t) = M(1 - 2e^{-t/\tau}) . \tag{26}$$

(a) *Initial Conditions*

Before solving for the equation of motion, a possible set of initial conditions must be established. For a PIPA operating in at least a 2:2 mode (minimum of two pulses in each direction under no input conditions) the float time constant, J/C , is small enough so that the velocity of the float at any switching time may be equated to M/C in magnitude. The initial position, $A_f(0)$, may be chosen to be at a point farther away from null than the minimum threshold angle that can be detected by the interrogator.

(b) *Solution*

Solving Equation (26) and eliminating second-order terms yields

$$A_f(t) = \frac{M}{C}\left[t - 2\tau - \frac{2J}{C}\left(1 - \frac{e^{-Ct/J}}{1 - \tau C/J}\right)\right] + A_f(0) . \tag{27}$$

4.2 Signal Generator

Equation (21) defined the signal generator output as

$$V_s(t) = -K_s I_p A_f(t) \sin \omega_0 t , \tag{28}$$

assuming the rate term is equal to zero. Substituting Equation (27) into Equation (28),

$$V_s(t) = -K_s I_p \left[\frac{M}{C} \left\{ t - 2\tau - \frac{2J}{C} \left(1 - \frac{e^{-Ct/J}}{1 - C\tau/J} \right) \right\} + A_f(0) \right] \sin \omega_0 t \ . \tag{29}$$

The tuned filter at the output of the SG has a transfer function

$$F(s) = \frac{(1/L_s C_s)}{s^2 + (R_T/L_s)s + (1/L_s C_s)} \ . \tag{30}$$

4.3 Total Equation

Combining Equations (29) and (30), and incorporating a gain factor to account for the amplification in the loop, the following relation is obtained:

$$V_I(s) = G V_s(s) F(s) \ , \tag{31}$$

where

$V_I(s)$	is the Laplace transform of the input to the interrogator
G	is the amplifier gain product
$V_s(s)$	is the Laplace transform of $V_s(t)$.

The solution to Equation (31) is presented in the Appendix. For the purpose of the present discussion the solution can be simplified to

$$V_I(t) = \frac{D}{C} \left[\frac{+ CA_f(0) - 2\tau M - (2MJ/C)}{F_1} \sin (\omega_0 t + 90^\circ) - \frac{2MF_2}{F_1^2} \cos (\omega_0 t + \tan^{-1} \alpha_0/\omega_0) + \right.$$
$$\left. + \frac{Mt \sin (\omega_0 t + 90^\circ)}{F_1} + E(t) \right] \ , \tag{32}$$

where the new terms introduced are defined as

D = constant multiplier resulting from fixed parameters in the loop

F_1, F_2 = constants derived from the SG secondary circuit parameters and the excitation frequency

$$\alpha_0 = \frac{\text{total resistance in SG secondary circuit}}{\text{2 SG secondary inductance}}$$

$E(t)$ = exponential terms in the equation .

The Apollo PIPA typically modes 2:2 or higher and, under these conditions, the exponential terms are negligible and will be omitted in the following discussion. Furthermore, if the time of interrogation is the same as the time of switching, $t = 2\pi n/\omega_0$, the value of $V_I(t)$ must be equal to or greater than the minimum threshold voltage of the interrogator, V_0 , for the given value of n , if switching is to occur. Thus, the requirement for switching may be expressed as:

$$V_0 \leqslant \frac{DM}{CF_1}\left[+\frac{C}{M}A_f(0) - 2\tau - \frac{2J}{C} - \frac{2F_2}{F_1}\cos\left(\omega_0 t + \tan^{-1}\alpha_0/\omega_0\right) + \frac{2\pi n}{\omega_0}\right]. \qquad (33)$$

Further simplifying,

$$0 \leqslant \overbrace{-C/M\left[\frac{V_0F_1}{D} - A_f(0)\right]}^{B_1} - \overset{B_2}{\frac{2J}{C}} + \overbrace{\left[\frac{2\pi n}{\omega_0} - 2\tau - \frac{2F_2}{F_1}\cos\frac{\omega_0}{(\alpha_0^2 + \omega_0^2)^{\frac{1}{2}}}\right]}^{B_3}. \qquad (34)$$

Equation (34) defines the dynamic response of the PIPA. B_1 represents the angle through which the float must move in order for a switching condition to occur. Both components of B_1 are of the same sign, since $A_f(0)$ is always negative relative to the direction of motion. For high values of C, B_1 predominates in the equation. B_2 represents the inertia effects due to reversing the momentum of the float. This term is the dominant factor in the equation for low values of C.

B_3 is the driving term, having to overcome the position and inertia terms. The magnitude of B_3 varies directly with n, the number of torque pulses applied in a given direction. The "-2τ" is present to account for the effect of the time constant in the torque current loop. The sinusoidal component represents an information lag due to the SG filter.

The interplay between the above terms determines the mode which will exist in a particular configuration and the response of the instrument to a given set of circumstances. The presence of an input acceleration modulates the driving term B_3 and results in an unbalanced pulse sequence, counteracting the acceleration input.

A further examination of Equation (34) indicates that, for a PIPA operating in a given test configuration, the only terms subject to variation in the equation are $A_f(0)$ and n. $A_f(0)$ will vary as a function of the interrogator sampling and the input acceleration. With no input, and with optimum sampling, $A_f(0)$ will be a minimum, and a minimum value of n will occur.

The minimum value of n can be shown to vary as a function of C by graphing the terms in Equation (34), with $A_f(0)$ equal to a minimum. In Figure 13, C is plotted as the abcissa and the terms B_1, B_2, and B_3 are plotted in absolute value along the ordinate. Dimensionally, each of the ordinate terms has the units of time.

B_3 is positive for $n = 1$ and is constant for any given value of n. The plot of B_3 thus appears as horizontal lines, differing in ordinate value by $2\pi\Delta n/\omega_0$ seconds. B_1 and B_2 appear as linear and hyperbolic plots, respectively.

In order for a switching condition to occur, Equation (34) must be positive. Thus, the minimum value of n for which switching will occur is the smallest value of n for which B_3 is greater than the sum of $B_1 + B_2$ (shown as a dashed line in Figure 13). For all C values, $B_3 < \Sigma(B_1 + B_2)$, for $n = 2$, and a 2:2 mode cannot exist. A 3:3 mode does exist for both C_1 and C_2 since, for these values of C, $(B_3)_{n=3} > \Sigma(B_1 + B_2)$.

It should be noted that the highest mode that will actually exist will be dependent on the maximum variation permissible in $A_f(0)$. $[\Delta A_f(0)]_{max}$ has a limiting value of $(M/C)\Delta T$, the steady-state distance the float will travel due to one torque pulse. In an operating instrument $[\Delta A_f(0)]_{max}$ is somewhat smaller than the limiting value, and the maximum variation in n will be less than one. For $C = C_2$ in Figure 13, the permissible range of $A_f(0)$ allows the sum of $B_1 + B_2$ to be greater than $(B_3)_{n=3}$, requiring the occurrence of a 4:4 mode. Thus, for operation at $C = C_2$, both 3:3 and 4:4 modes will exist. For $C = C_1$, $(B_3)_{n=3}$ is always greater than the sum of $B_1 + B_2$ and only the 3:3 mode exists.

4.4 Computer Simulation Results

The complete form of Equation (34) (Equation (A3) in the Appendix) was programmed on an IBM 360 computer and used to determine the average moding of the PIPA as a function of C, J, and M. The results are presented in Figures 14-16.

The curve obtained for varying C in Figure 14 basically resembles the sum of terms $B_1 + B_2$ shown in Figure 13. In the case of the actual unit parameters, the $2J/C$ decay closely approximated the rise of the linear function of C in the region near intercept, resulting in a marked flattening of the curve.

Figure 15 shows that the PIPA average mode increases as J increases. Increasing J amplifies the inertia effects and increases the time required for the float to traverse the angle between interrogator thresholds, resulting in a higher mode.

A plot of average mode as a function of M is shown in Figure 16. For small values of M, B_1 in Equation (34) becomes large relative to the distance the float travels due to one torque pulse, causing the moding to increase. For large values of M, B_1 is negligible compared to B_2 and B_3, and an essentially constant mode is obtained, independent of M.

Figures 17 and 18 show the computer-predicted response of the PIPA to a square wave input over a wide frequency range. The input was purposely unbalanced so that a net input would result. In this way a required number of ΔV's (restoring pulses) to be expected from a perfect accelerometer could be calculated and compared with the output predicted by the computer.

Figure 17 shows the computed outputs for a CM PIPA to be within less than 1% of the actual input, up to a frequency of 1500 Hz. Figure 18 shows similar results for a LM PIPA. These results are not surprising when it is recalled that an implicit assumption in the design of the PIPA is that the float acts as a perfect integrator. Figures 17 and 18 bear out that assumption.

5. OPERATIONAL INSTABILITIES

There are three parameters of importance in determining the stability and performance of the PIPA: scale factor, bias, and input-axis position.

5.1 Scale Factor and Bias Equations

The scale factor, as defined in Section 2, was given as

$$SF = \frac{(M^{+} + M^{-})\Delta T}{2ml} . \tag{35}$$

The torque M produced by the torque generator may be written in terms of the microdyn parameters as

$$M = 2N^{2} r d \mu I^{2} k / g_{n} . \tag{36}$$

where

M	is the electromagnetic torque applied in either the positive or negative direction
N	is the number of turns per coil pair
r	is the outside radius of the rotor
d	is the effective length of the rotor
I	is the torque current
μ	is the permeability of free space
g_n	is the nominal air gap
k	is a correction factor to account for core losses in the iron.

The product, $N^2 r d \mu$, may be equated to some constant, K_2, and Equation (36) substituted into Equation (35) to give

$$SF = \frac{K_2 I^2 \Delta T (k^+ + k^-)}{m \ell g_n} . \tag{37}$$

The correction terms k^+ and k^-, referring to the positive and negative torquing directions, will in general not be equal, since the magnetic states of the iron in the two torquer flux paths will normally differ. In order to be completely general, the ΔT term should indicate the effects of different time constants in the torquer compensation circuits and possible inequalities in the switching characteristics of the output transistors in the binary current switch. For the purpose of the present discussion it is assumed that these terms have been adequately compensated such that a nominal value of ΔT can be used for both positive and negative torquing.

The bias expression corresponding to the expression for scale factor shown in Equation (37) is

$$\text{bias} = \frac{K_2 I^2 \Delta T}{m \ell} \left[\frac{k^+}{g^+} - \frac{k^-}{g^-} \right] . \tag{38}$$

The air gaps in general will not be equal, and thus g^+ and g^- represent the effective air gaps for positive and negative torquing respectively. Equation (38) is written assuming that the bias compensation network, shown in Figure 11, is open. The addition of this circuit will predictably alter the discussion presented here and, for clarity, it is omitted.

5.2 Variation in I

The stability of the torquer current is of major importance in maintaining a stable scale factor. The degree of current stability will in great measure depend on the stability of the PVR and the scale factor resistor. The PVR is a cascaded zener diode voltage source. The final stage is chosen to operate at a voltage where the temperature coefficient of the diode crosses through zero. Voltage deviations of the PVR are less than 10 ppm/year, with a temperature coefficient of less than 3 ppm/°C.

The scale factor resistor is a parallel combination of precision resistors with low temperature coefficients. One leg of the network contains a potentiometer which allows accurate adjustment of the current to any desired value. Resistance stability is approximately 3 ppm/year.

The d.c. amplifier is designed to keep transients as small as possible and to minimize current changes due to variations in supply voltage.

From Equation (25) it is evident that variations in current will have little effect on bias stability unless $\{(k^+/g^+) - (k^-/g^-)\}$ is large. Since the bias is kept at a small value, the primary effect will be on the scale factor of the instrument.

5.3 Variations in k^+ and k^-

The primary source of scale factor and bias instability in the PIPA is the variation in the magnetic state of the iron in the torquing circuit. The magnetic parameters of the iron are affected by temperature changes, large float rotations, torquer current overshoots, and variations in 3200-Hz excitation voltage to the magnetic suspensions. Temperature variations cause changes in stress exerted by the potting encapsulating the torquer microsyn. The changes in stress cause corresponding changes in permeability of the iron. Since the effect of the potting is similar for both positive and negative torquing circuits, temperature changes primarily affect scale factor, and the bias changes are almost negligible.

A change in bias and scale factor can occur if a malfunction in the torquing loop causes an unstable condition long enough to torque the float to a rotary stop. In such a case, full flux density will exist along portions of the rotor which normally see only fringing fields, resulting in a shift in magnetic state. The effect will be primarily a bias effect, since one direction of torquing will be predominantly affected.

Torque current overshoots may result from shorts in the circuit or a large overvoltage applied to the d.c. current source. The effect of an overshoot is to drive the torque generator microsyn to a higher operating point on the B-H curve. Upon removal of the overshoot condition the torquer will return to an operating point somewhat higher than the initial one, with a resulting increase in M. If the overshoot occurred long enough to affect both torquers, the primary change seen will be in scale factor. If only one torquer is affected, a large bias shift will result.

The single-piece rotor on the TG end of the PIP carries flux from both the suspension and TG circuits. Variations in the level of the suspension flux can cause variations in k^+ and k^-. These terms are particularly sensitive to suspension overshoots which could drive portions of the rotor into saturation. Depending on the geometry involved, either bias or scale factor shifts could result.

Good design practices to minimize the possibility of overshoots and torquing loop instabilities can eliminate a major portion of the instability in k^+ and k^-. Adequate control of temperature and excitation voltage can maintain the other effects to within acceptable limits.

5.4 Position Memory

Position memory is an acceleration-sensitive bias exhibited by the PIPA, the cause of which is still under investigation. Position memory causes a bias change of ± 0.1 cm/sec^2 for a change in input to the accelerometer from 0 to ± 1g. The change in position memory bias as a function of input is predictable and thus may be compensated for if the expected acceleration profile is known.

5.5 A.C. Hysteresis

The term "a.c. hysteresis" is applied to bias variations caused by suspension fluxes coupling through fringe portions of the rotor when the float is allowed to rotate to an angular stop. This effect can be controlled by adequate specification of the magnetic properties of the rotor material.

5.6 Pendulosity Changes

The hollow beryllium float will exhibit changes in pendulosity if damping fluid is allowed to seep into any portion of the float. Proper design will virtually eliminate any possibility of such an occurrence. If such a condition were to occur, an input axis shift would be the first effect noted, since the sensing direction for input acceleration would change relative to the float body. Except in extreme cases, effects on scale factor and bias would be negligible.

5.7 Changes in Air Gap

Changes in g^+ and g^- will result if the magnetic suspension allows motion of the float within the case of the instrument. For the most part, the suspension parameters are quite stable and negligible motion of the float will occur. (Under conditions of high input acceleration along the output axis of the instrument, the suspension will be required to counteract the torque produced by the output axis acceleration acting on the pendulosity of the instrument, resulting in some float motion.) If a malfunction should occur, large bias shifts will normally result, since the differential term in Equation (38) is very sensitive to changes in g. There will also be some change in scale factor, but it will be limited, since scale factor is primarily a function of g_n.

5.8 Fluid Impurities

Contaminants or gas bubbles within the damping fluid can become positioned between the float and the case of the PIPA and exert error torques on the float. Proper filling techniques will minimize the possibility of foreign particles in the fluid at the time of build. Gas bubbles can result from air leaking into the instrument from outside or from helium gas leaking into the fluid from the hollow float or from the bellows.

A cemented metal band is used to seal the joint between the two end housings of the PIP and this joint, as well as the float and bellows seals, are carefully checked before filling of the unit with damping fluid. Filling of the instrument is accomplished through small fill holes at either end of the case. After fill, these holes are sealed using metallic compression seals.

By use of the above technique, the possibility of air leaking into the instrument is virtually eliminated. Slow leaking from the float can easily be kept to acceptable levels with presently known cementing and fabrication techniques.

The bellows remain one of the most critical elements in the design. Rigid screening tests on the bellows, and careful inspection by X-ray techniques after completion of build, will reduce bellows failures to a minimum.

6. PERFORMANCE CAPABILITIES AND POSSIBLE IMPROVEMENTS

6.1 Performance

The discussion in Section 5 dealing with the PIPA instabilities indicates that, to a great extent, the performance capability of the instrument is dependent upon how well its environment is controlled. With the exception of position memory and a.c. hysteresis, the magnetic instabilities can be kept small by good control of temperature, torquing current and 3200-Hz excitation voltage.

Scale factor change due to temperature variation is one of the major limitations of the present instrument. The CM PIPA exhibits approximately 360 ppm change in scale factor per °C. The LM PIPA exhibits approximately 630 ppm change per °C.

The performance table shown in Figure 19 assumes a temperature stability of ± 0.14°C and no magnetic shifts due to abnormal current and voltage levels. The bias stabilities quoted are primarily a function of position memory and a.c. hysteresis.

6.2 Possible Improvements

Use of a potting compound for the TG microsyn, which would decrease shrinkage stresses by a factor of two, would reduce scale factor sensitivity to temperature by at least an order of magnitude. Use of a three-piece rotor to separate the suspension and TG fluxes would eliminate the coupling between the suspension and TG circuits. The three-piece rotor, coupled with optimum rotor design, would also reduce the effects of a.c. hysteresis.

With the incorporation of improvements in the above areas, the PIPA performance could be substantially improved over present Apollo requirements, and the instrument could be made suitable for longer and more demanding missions.

ACKNOWLEDGEMENT

This paper was prepared under DSR Project 55-23870, sponsored by the Manned Spacecraft Center of the National Aeronautics and Space Administration through Contract NAS 9-4065 with the Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts.

Its publication does not constitute approval by the National Aeronautics and Space Administration of the findings or the conclusions contained therein. It is published only for the exchange and stimulation of ideas.

REFERENCES

1. Gilinson, P.J. Jr et al. — *A Magnetic Support for Floated Inertial Instruments*. Report No. R-277 Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, April 1960.

2. Wiener, Thomas Freud — *Theoretical Analysis of Gimballess Inertial Reference Equipment Using Delta-Modulated Instruments*. ScD Thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts.

3. Miller, J.E. McNeil, J. — *The PIPA (Pulsed Integrating Pendulum Accelerometer)*. E-934, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, May 1960.

4. Sitomer, J.L. Lynch, P.J. — *Apollo Block II - PIPA Electronics and Gyro Pulse Torque Electronics*. APM No.904, Instrumentation Laboratory, Massachusetts Institute of Technology, Cambridge, Massachusetts, March 1964.

APPENDIX

Starting with Equation (31) in the text,

$$V_I(s) = GV_s(s)F(s) \; , \qquad \text{(A1)}$$

$V_s(s)$ is first obtained by transforming Equation (29). The result is multiplied by $F(s)$ as defined in Equation (30) with the result

$$V_I(s) = \frac{-GK_sI_pM}{CL_sC_s}\left[\frac{1}{s^2+\frac{R_I}{L_s}s+\frac{1}{L_sC_s}}\right]\left[\left(-2\tau-\frac{2J}{C}+\frac{C}{M}A_f(0)\right)\frac{\omega_0}{s^2+\omega_0^2}+\right.$$

$$\left. + \frac{2\omega_0 s}{(s^2+\omega_0^2)} + \frac{2J^2}{C(J-\tau C)}\,\frac{\omega_0}{(s+C/J)^2+\omega_0^2}\right] . \qquad \text{(A2)}$$

Taking the inverse transform of Equation (A2),

$$V_I(t) = -\frac{GK_sI_p}{CL_sC_s}\left\{\frac{(-2\tau M - 2JM/C + CA_f(0))}{[(\lambda_0^2+\omega_1^2-\omega_0^2)^2+4\lambda_0^2\omega_0^2]^{\frac{1}{2}}}\left[\sin\left(\omega_0 t - \tan^{-1}\frac{2\lambda_0\omega_0}{\lambda_0^2+\omega_1^2-\omega_0^2}\right)+\right.\right.$$

$$\left. + \frac{\omega_0}{\omega_1}e^{-\lambda_0 t}\sin\left(\omega_1 t + \tan^{-1}\frac{2\lambda_0\omega_1}{\lambda_0^2-\omega_1^2+\omega_0^2}\right)\right] -$$

$$- \frac{2(M)(\omega_0^2+\lambda_0^2)^{\frac{1}{2}}}{(\lambda_0^2+\omega_1^2-\omega_0^2)^2+4\lambda_0^2\omega_0^2}\cos\left(\omega_0 t + \tan^{-1}\frac{\lambda_0}{\omega_0} + 2\tan^{-1}\frac{2\lambda_0\omega_0}{\lambda_0^2+\omega_1^2-\omega_0^2}\right) +$$

$$+ \frac{(M)(t)\sin\left(\omega_0 t - \tan^{-1}\frac{2\lambda_0\omega_0}{\lambda_0^2-\omega_0^2+\omega_1^2}\right)}{[(\lambda_0^2-\omega_0^2+\omega_1^2)^2+4\omega_0^2\lambda_0^2]^{\frac{1}{2}}} +$$

$$+ \left(\frac{2J^2M}{C(J-\tau C)}\right)\left[\frac{e^{-Ct/J}\sin\left(\omega_0 t - \tan^{-1}\frac{2\omega_0(\lambda_0 - C/J)}{(\lambda_0 - C/J)^2-\omega_1^2+\omega_0^2}\right)}{([(\lambda_0 - C/J)^2-\omega_1^2+\omega_0^2]^2+4\omega_0^2(\lambda_0 - C/J)^2)^{\frac{1}{2}}} +\right.$$

$$\left.\left. + \frac{\frac{\omega_0}{\omega_1}e^{-\lambda_0 t}\sin\left(\omega_1 t - \tan^{-1}\frac{2\omega_1(C/J-\lambda_0)}{(C/J-\lambda_0)^2-\omega_1^2+\omega_0^2}\right)}{([(C/J-\lambda_0)^2-\omega_1^2+\omega_0^2]^2+4\omega_1^2(C/J-\lambda_0)^2)^{\frac{1}{2}}}\right]\right\} . \qquad \text{(A3)}$$

where

$$\omega_1 = \left(1/L_sC_s - \frac{R_I^2}{4L_s^2}\right)^{\frac{1}{2}}$$

$$\alpha_0 = R_T/2L_s$$

$$\omega_0 = \text{excitation frequency} .$$

Noting that, for the tuned filter, $\alpha_0^2 + \omega_1^2 - \omega_0^2 = 0$,

$$\tan^{-1} \frac{2\alpha_1\omega_0}{\alpha_0^2 + \omega_1^2 - \omega_0^2} = 90^\circ .$$

Furthermore, since the exponential terms are small and are negligible for most applications, they will be grouped separately. Thus Equation (A3) can be simplified to

$$V_I(t) = \frac{GK_sI_p}{CL_sC_s}\left[\frac{+CA_f(0) - 2\tau M - \frac{2MJ}{C}}{2\alpha_0\omega_0}\sin(\omega_0 t + 90^\circ) - \right.$$

$$- \frac{2M(\omega_0^2 + \alpha_0^2)^{\frac{1}{2}}}{4\alpha_0^2\omega_0^2}\cos(\omega_0 t + \tan^{-1}\alpha_0/\omega_0) +$$

$$\left. + \frac{Mt\sin(\omega_0 t + 90^\circ)}{2\alpha_0\omega_0} + E(t)\right] . \tag{A4}$$

where $E(t)$ represents the exponential terms.

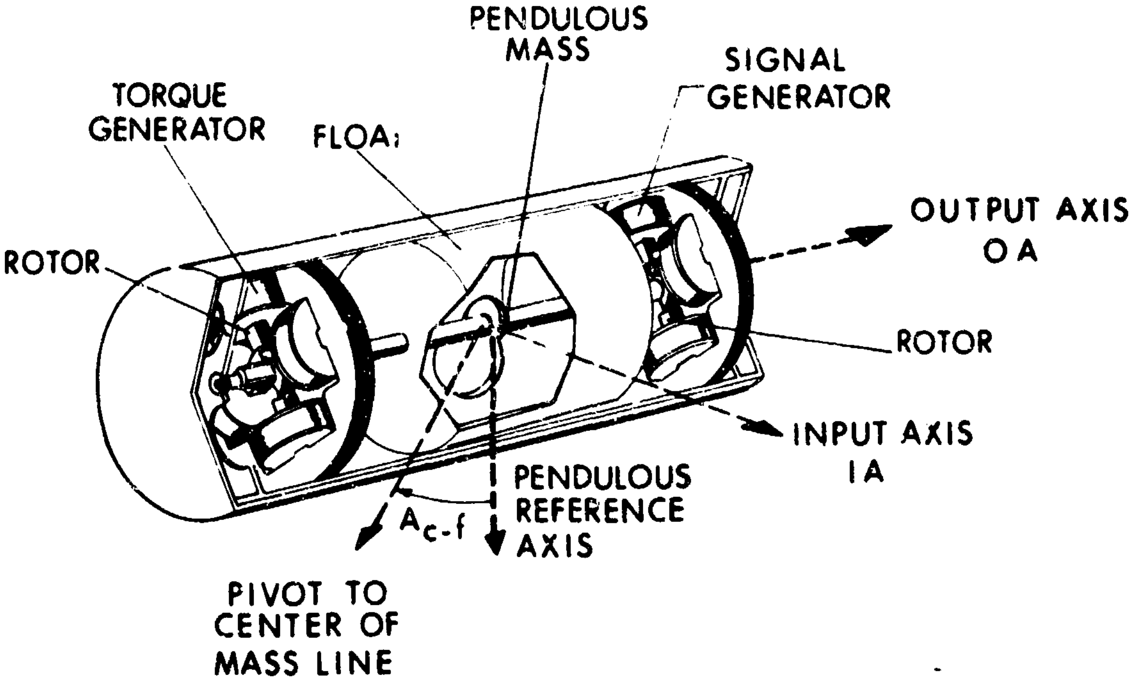


Fig. 1 Pulsed Integrating Pendulum schematic

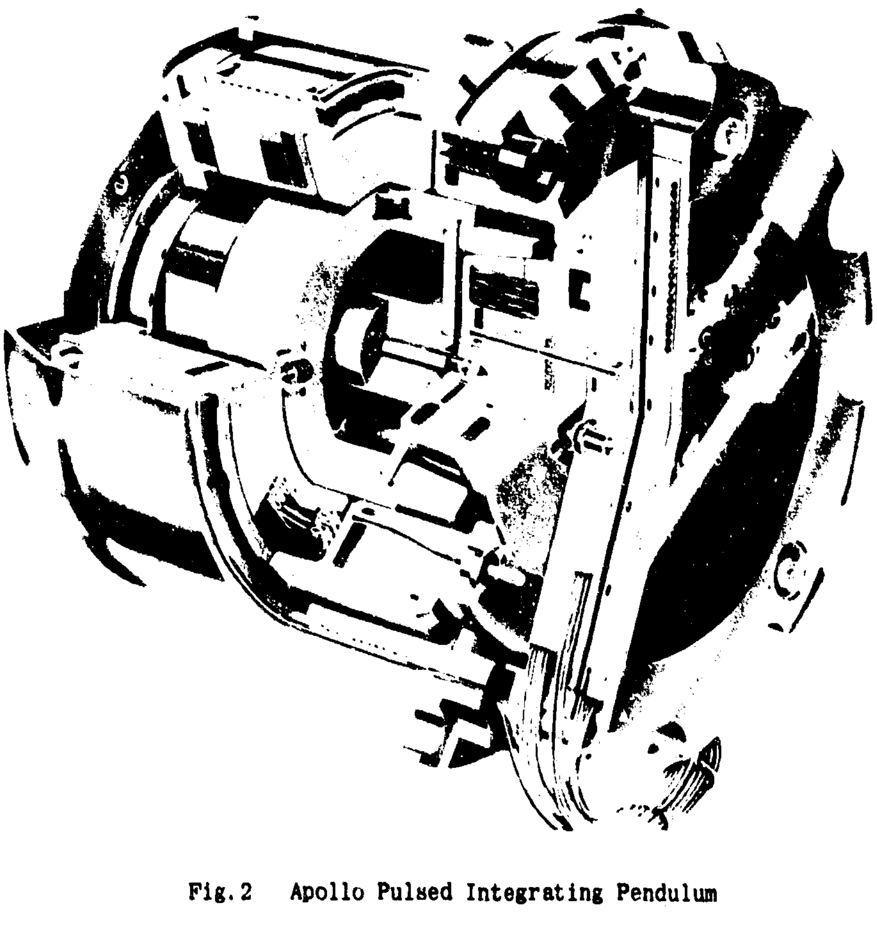
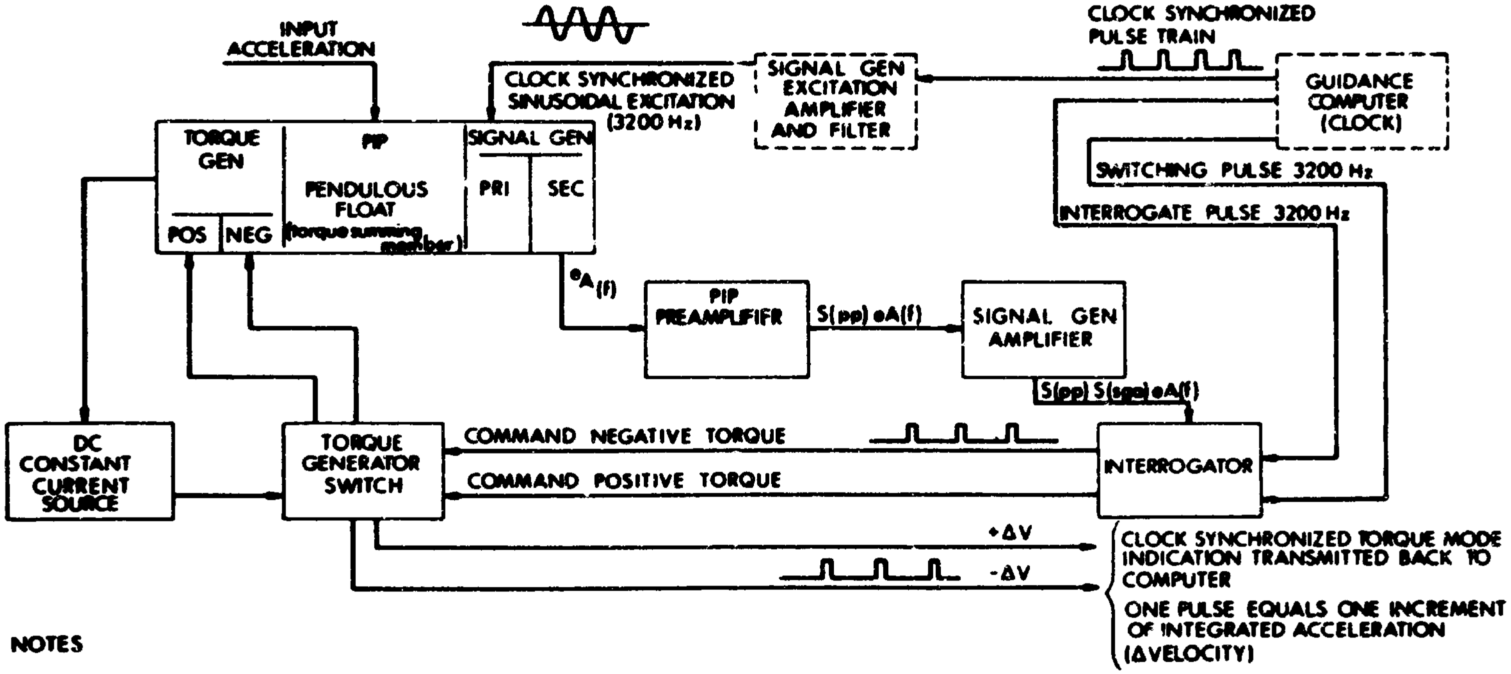


Fig. 2 Apollo Pulsed Integrating Pendulum


INPUT ACCELERATION
TORQUE GEN
POS
NEG
PIP
PENDULOUS FLOAT
(torque summing member)
SIGNAL GEN
PRI
SEC
CLOCK SYNCHRONIZED SINUSOIDAL EXCITATION (3200 Hz)
SIGNAL GEN EXCITATION AMPLIFIER AND FILTER
CLOCK SYNCHRONIZED PULSE TRAIN
GUIDANCE COMPUTER (CLOCK)
SWITCHING PULSE 3200 Hz
INTERROGATE PULSE 3200 Hz
eA(f)
PIP PREAMPLIFIER
S(pp) eA(f)
SIGNAL GEN AMPLIFIER
S(pp) S(sga) eA(f)
DC CONSTANT CURRENT SOURCE
TORQUE GENERATOR SWITCH
COMMAND NEGATIVE TORQUE
COMMAND POSITIVE TORQUE
INTERROGATOR
+ΔV
-ΔV
CLOCK SYNCHRONIZED TORQUE MODE INDICATION TRANSMITTED BACK TO COMPUTER
ONE PULSE EQUALS ONE INCREMENT OF INTEGRATED ACCELERATION (ΔVELOCITY)

NOTES

1. PIP (PULSED INTEGRATING PENDULUM) SINGLE DEGREE OF FREEDOM PENDULOUS UNIT
2. WAVE FORMS SHOW CONDITION FOR NEGATIVE INPUT ACCELERATION OF PIP CASE

Fig.3 Apollo PIPA block diagram

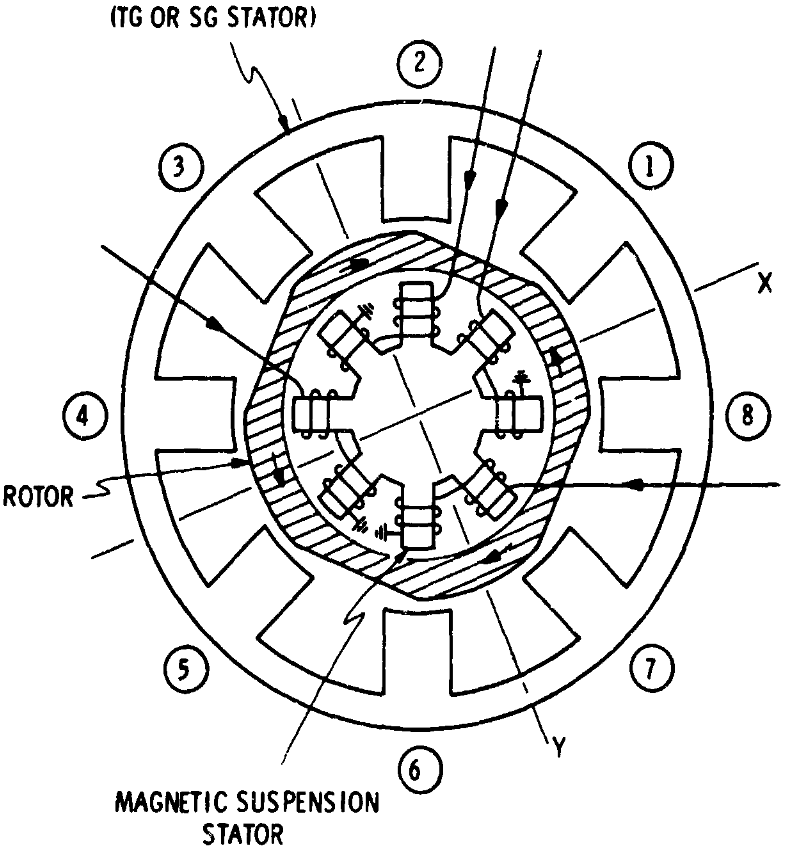


Fig.4 Apollo PIP magnetic suspension cross-section

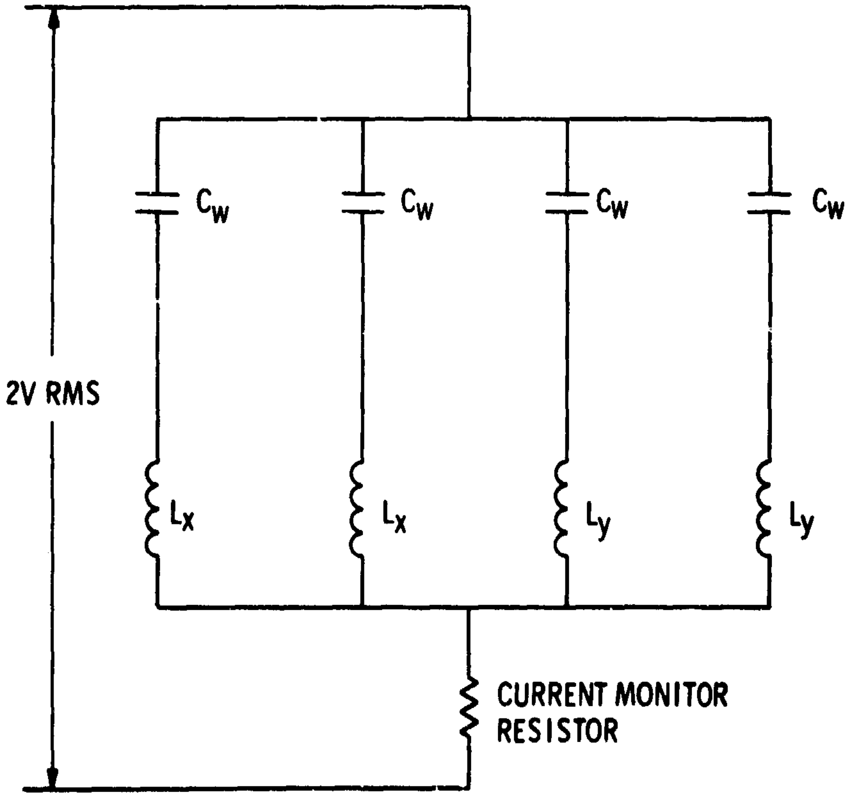


Fig.5 Apollo PIP magnetic suspension circuit

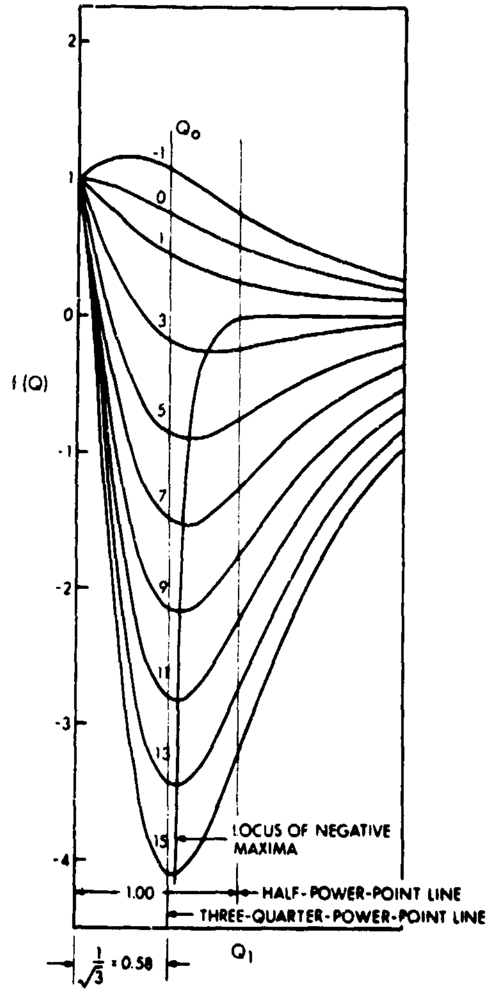


Fig.6 Locus of maxima and minima of f(Q), shown to enlarged scale for $0 < Q_1 < 2$

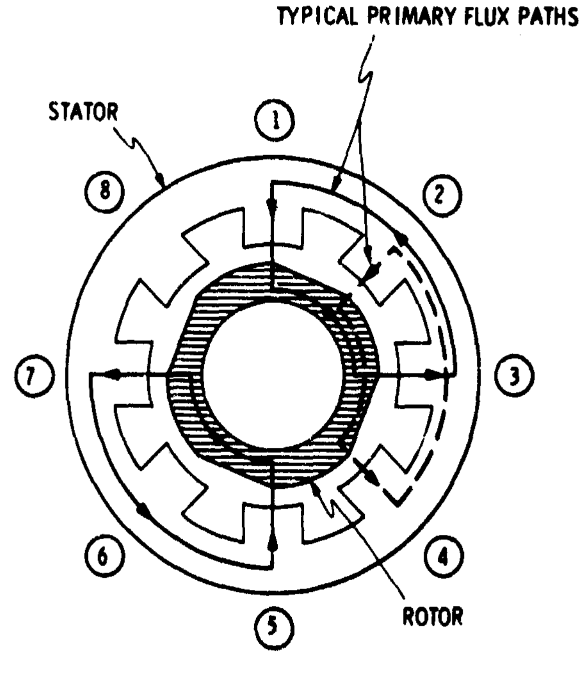


Fig.7 Apollo PIP signal generator cross-section

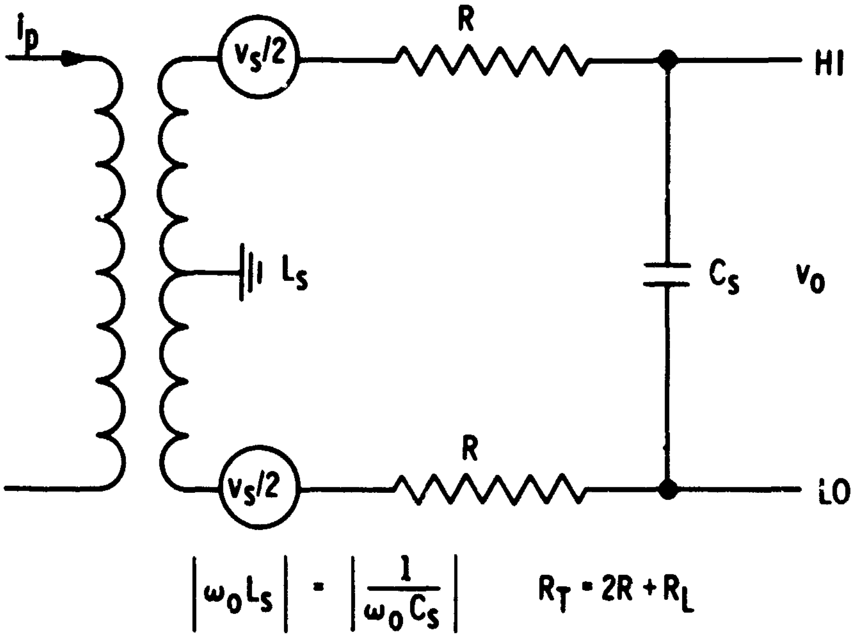


Fig.8(a) Apollo PIP SG secondary circuit

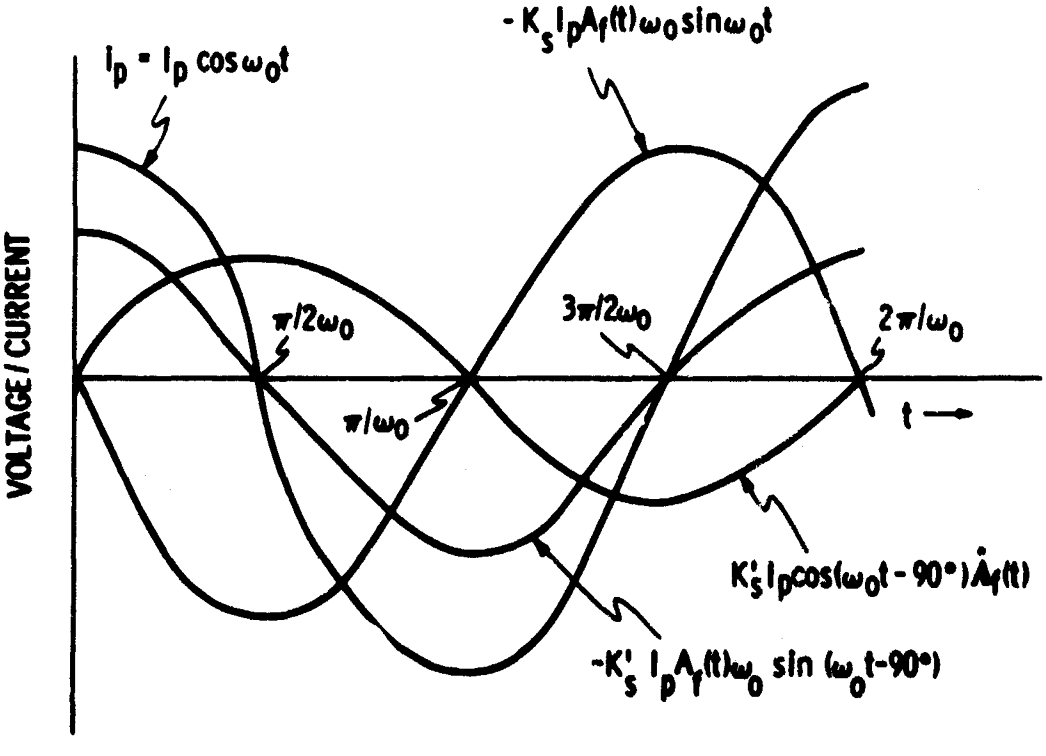


Fig.8(b) Apollo PIP SG signal phase relationships

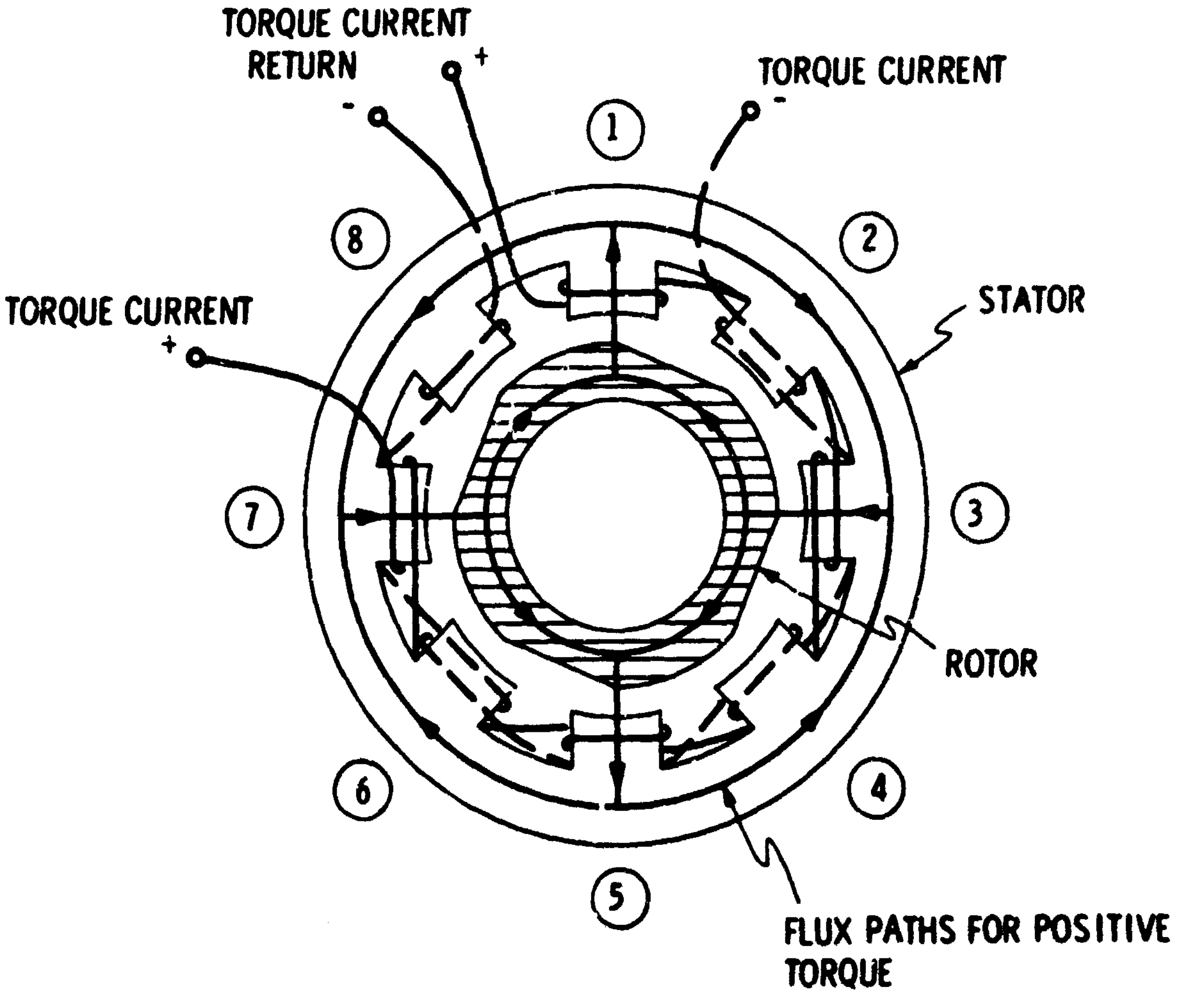


Fig.9 Apollo PIP torque generator cross-section

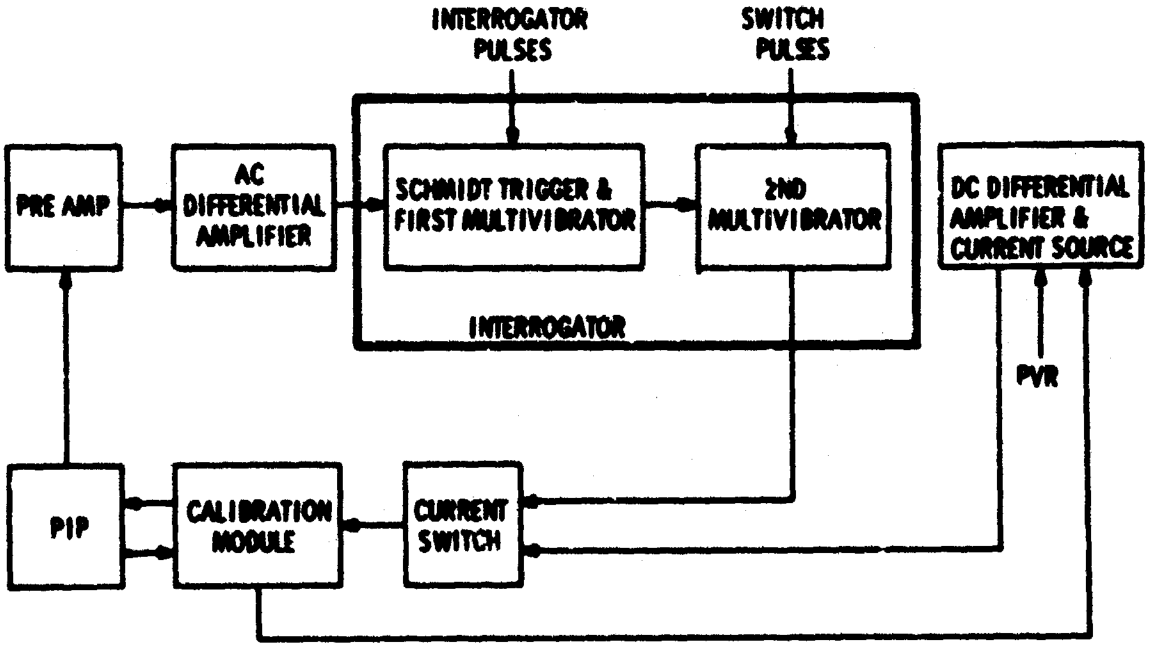


Fig.10 Apollo PIPA electronics control loop

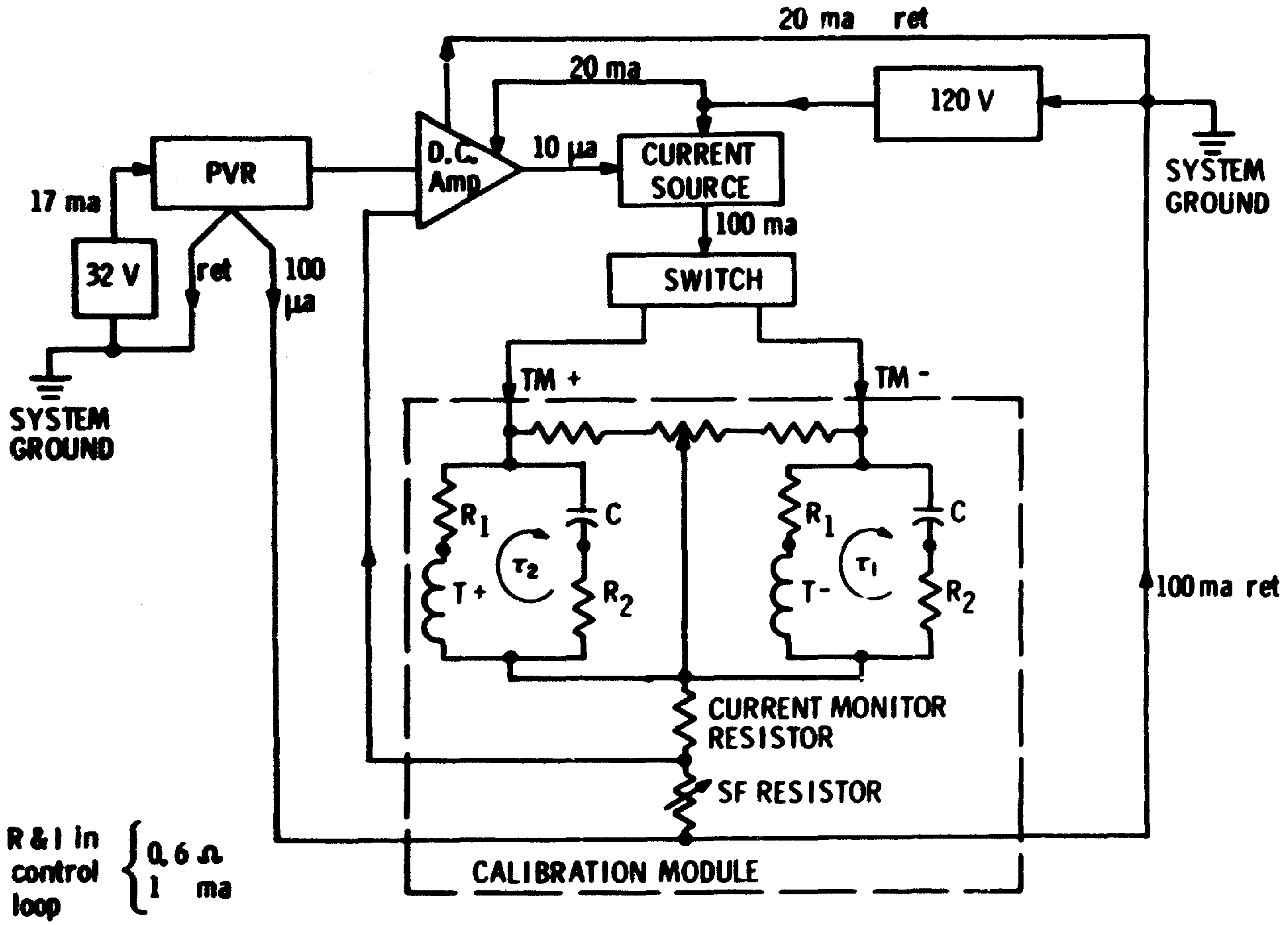


Fig.11 Apollo PIPA d.c. current loop

PARAMETER	MAGNITUDE
ANGULAR MOMENT OF INERTIA (J)	13 gm $\overline{cm}^2$
DAMPING COEFFICIENT (C)	120,000 dyne cm sec/rad
PENDULOSITY (ml)	0.25 gm cm
TORQUE CONSTANT (M)	
LM	800 dyne cm
CM	4680 dyne cm
TORQUE CURRENT (I_{TM})	
LM	43 ma
CM	103 ma
SCALE FACTOR (SF)	
LM	1.0 cm/sec pulse
CM	5.85 cm/sec pulse
OPERATING TEMPERATURE	54° C
EXCITATION VOLTAGE (3200Hz)	2.0 V rms
SUSPENSION CURRENT	70 ma (per suspension)
SUSPENSION STIFFNESS	1160 dynes/micron
SG SENSITIVITY	10 mv/mr rms
ELASTIC RESTRAINT	<0.1 dyne/mr
POWER TO PIP (OPEN LOOP)	0.25 watts
POWER TO PIP (CLOSED LOOP)	0.75 watts
POWER (ELECTRONICS)	12 watts
INTERROGATION FREQUENCY (ω_0)	3200 Hz

Fig.12 Apollo PIPA operating parameters

@ $C = C_1$, 3:3 MODE EXISTS

@ $C = C_2$, 3:3 + 4:4 MODES EXIST

@ ALL C VALUES, 2:2 MODE CANNOT EXIST

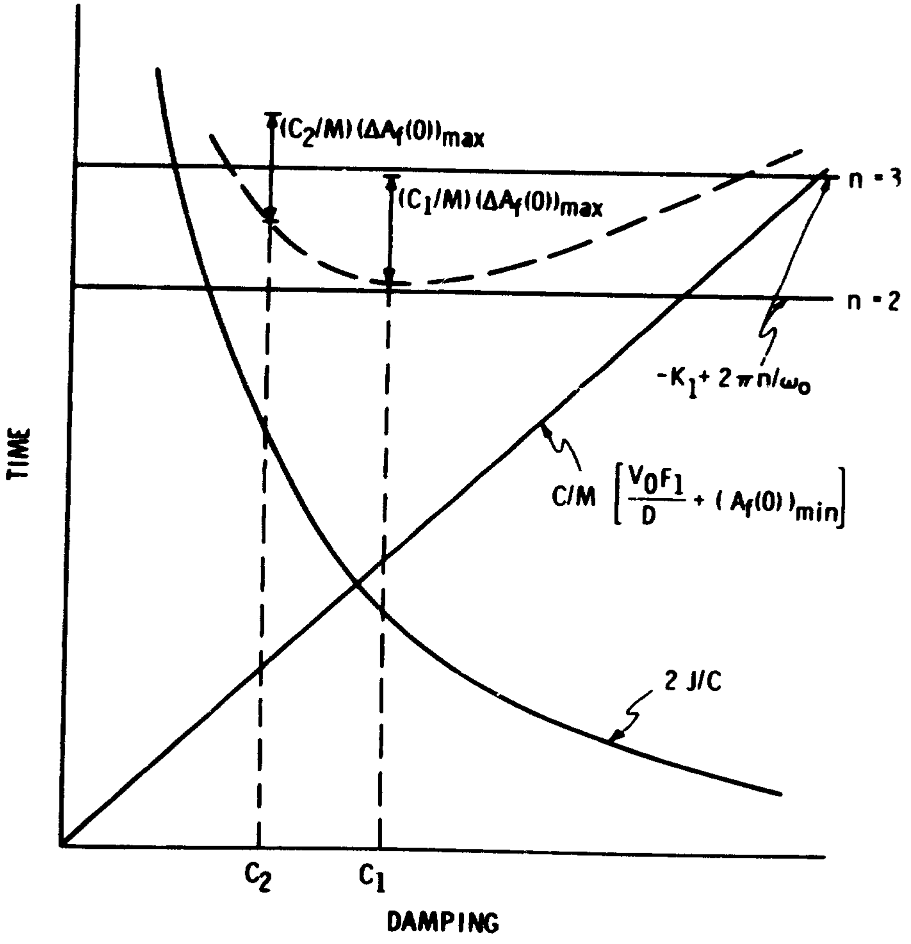


Fig.13 Graph of terms in Equation (34), showing possible modes of operation

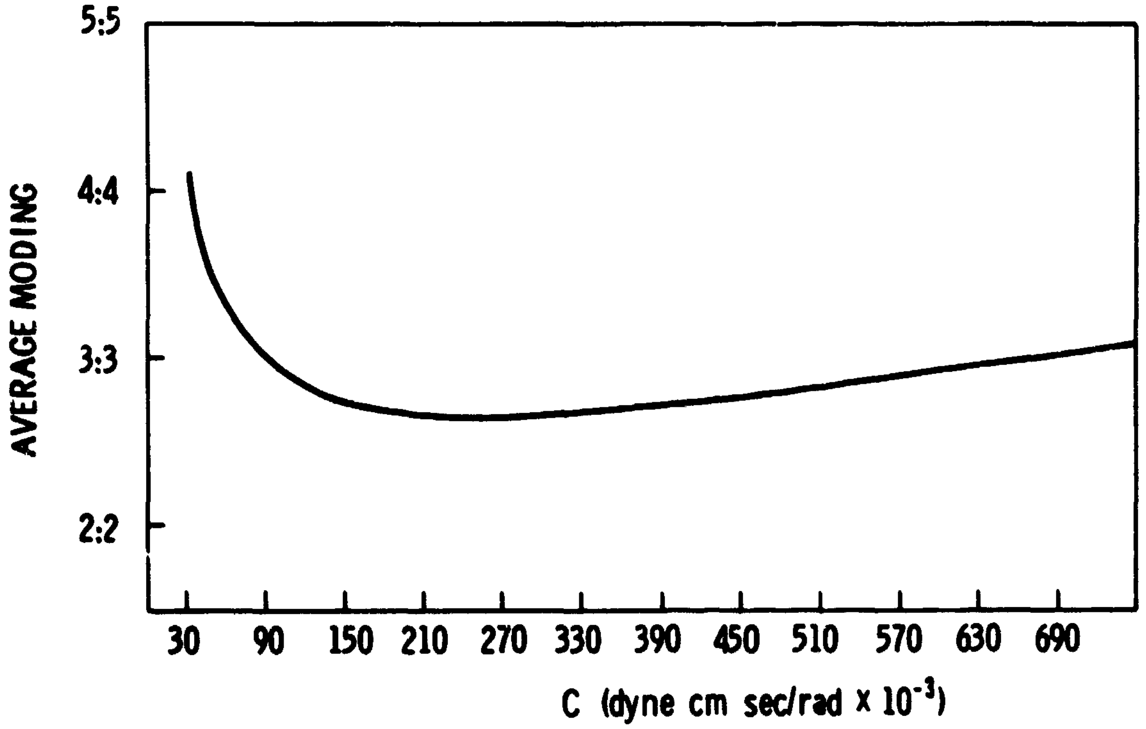


Fig. 14 Average moding versus damping, CM PIPA

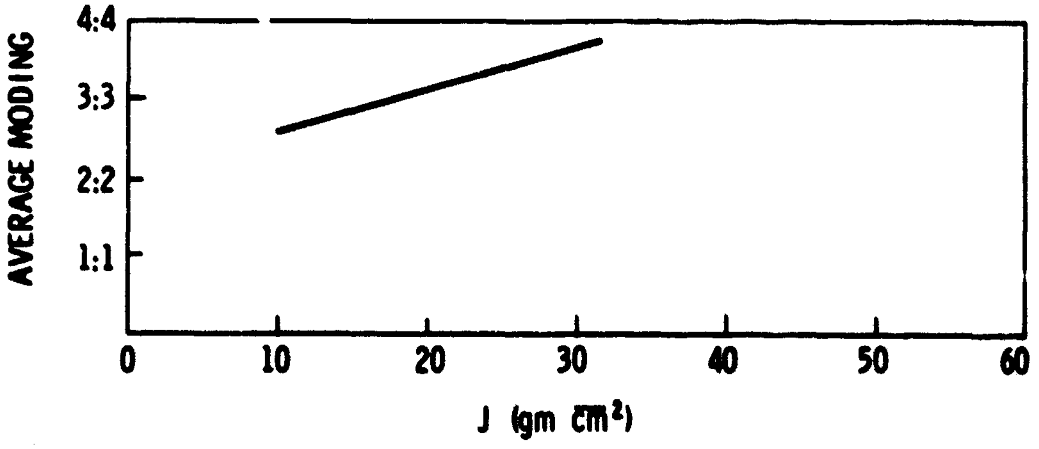


Fig. 15 Average moding versus angular moment of inertia, CM PIPA

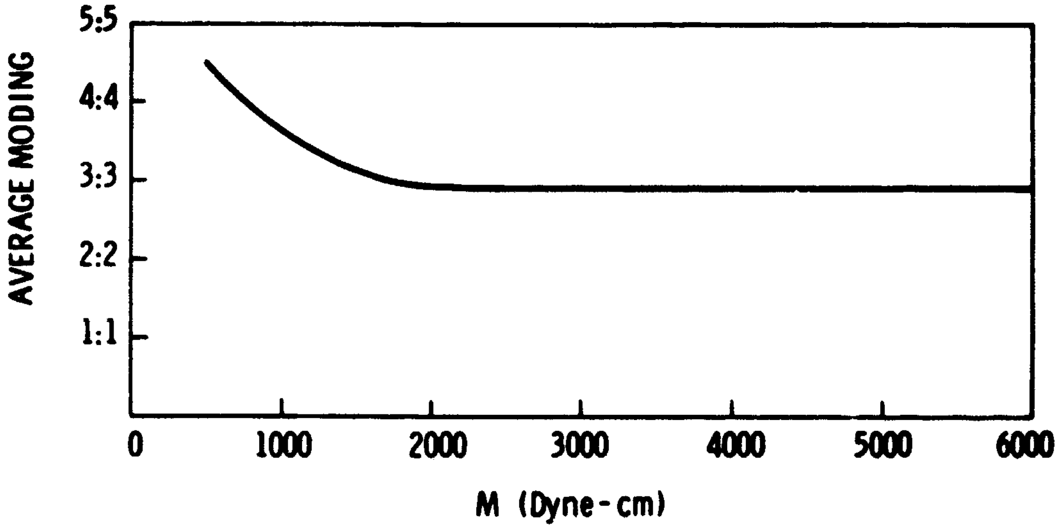


Fig. 16 Average moding versus torque constant, Apollo PIPA

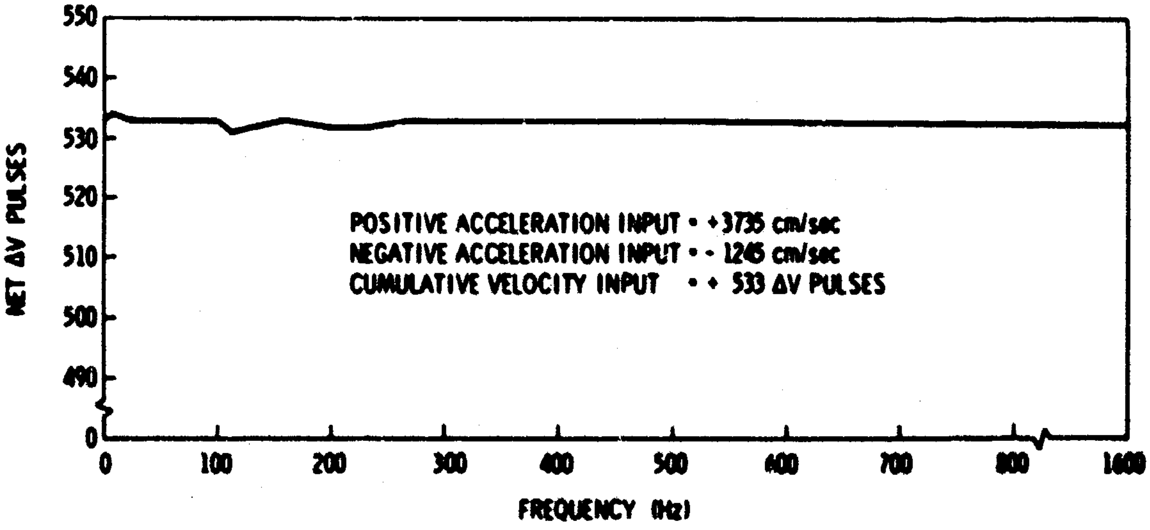


Fig. 17 Computer simulation results. CM PIPA frequency response to unbalanced square wave input

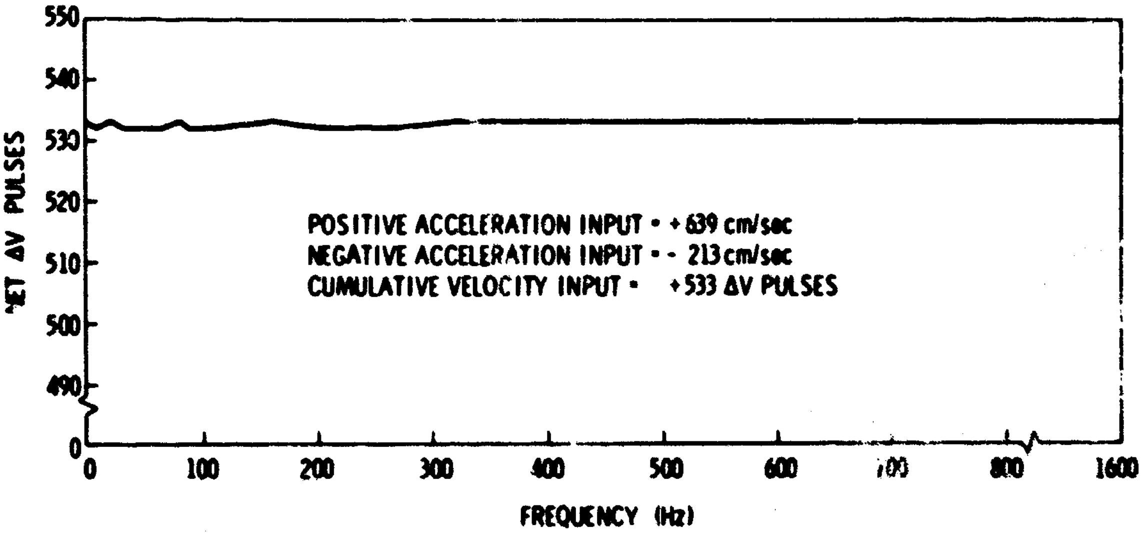


Fig. 18 Computer simulation results. LM PIPA frequency response to unbalanced square wave input

	CM	LM
SCALE FACTOR STABILITY	± 100 ppm	± 150 ppm
BIAS STABILITY	± 0.25 cm/sec^2	± 0.3 cm/sec^2
IA STABILITY	± 0.05 mr	± 0.05 mr
NON-LINEARITY	< 10 ug/g^2	< 10 ug/g^2
OPERATING MODE	2:2 + 3:3	3:3
RESOLUTION (MINIMUM THRESHOLD)	< 0.05 cm/sec^2	< 0.05 cm/sec^2

Fig. 19 Apollo PIPA closed-loop performance

L'ACCELEROMETRE ONERA A GRANDE SENSIBILITE

par

Michel Delattre

Office National d'Etudes et de Recherches Aérospatiales
29 Avenue de la Division Leclerc, 92 Chatillon, France

L'ACCELEROMETRE ONERA A GRANDE SENSIBILITE

Michel Delattre

L'accéléromètre ONERA à grande sensibilité (seuil de mesure $10^{-8} m/s^2$) a fait l'objet de plusieurs publications[1-4].

Références 1 et 2 traitent plus particulièrement de la théorie de cet accéléromètre. Référence 4 cite les différentes recherches actuellement en cours sur des accéléromètres très sensibles et analyse les problèmes soulevés par le développement de l'accéléromètre ONERA.

La présente publication a pour objet de décrire l'état actuel de ce développement.

1. RAPPEL DU PRINCIPE

Considérons une enceinte C disposant d'un degré de liberté suivant l'axe XX' (voir Figure 1). A l'intérieur de cette enceinte, sur l'axe XX' plaçons une masse m soumise à une force de rappel élastique $-Kx$ (voir Figure 1) dès qu'elle s'éloigne de sa position d'équilibre 0' au centre du cadre. Supposons qu'il existe une force d'amortissement de m proportionnelle à sa vitesse de déplacement par rapport à 0', $-f(dx/dt)$.

Supposons que pour $t = 0$ le point 0' se trouve en 0 et que le cadre subisse une accélération Γ_x (x est toujours de signe contraire à Γ_x).

A l'instant t, si l'on considère l'équilibre des forces appliquées sur m, on peut écrire

$$m \frac{d^2(X - x)}{dt^2} - Kx - f \frac{dx}{dt} = 0$$

$$\Gamma_x = \frac{d^2X}{dt^2} = \frac{d^2x}{dt^2} + \frac{f}{m}\frac{dx}{dt} + \frac{K}{m} x .$$

Pour une accélération constante Γ_x, après évanouissement du phénomène transitoire, on a

$$\Gamma_x = \frac{K}{m} x . \qquad (1)$$

x est une mesure de Γ_x.

Avec un système disposant de trois degrés de liberté suivant les trois axes d'un trièdre tri-rectangle il est possible de connaître à la fois l'amplitude et la direction d'une accélération par trois mesures de déplacement.

2. DESCRIPTION SOMMAIRE

La masse d'épreuve est une sphère de béryllium de 40 mm de diamètre (poids 60 gr) contenue à l'intérieur d'une cage également sphérique.

La position de la bille par rapport à la cage est détectée suivant trois axes trirectangles passant par le centre de la cage.

Les détecteurs de position de la sphère sont du type à capacitance variable (voir Figure 2). Ils sont constitués par des électrodes dites "électrodes de détection" en forme de calottes sphériques (voir Figure 3).

Les forces de rappel, de nature électrostatique, sont imposées par l'intermédiaire d'électrodes dites "électrodes d'action" de forme annulaire sphérique.

Le principe de la chaîne électronique est montré sur la Figure 4. Les capacités "bille-électrode de détection" sont connectées à un pont capacitif alimenté par une tension de fréquence 500 kHz.

Les mouvements de la sphère font apparaître un signal de déséquilibre du pont.

Après amplification, ce signal v s'ajoute ou se retranche de la tension de polarisation V_0 appliquée en permanence sur les électrodes d'action. Une électrode d'action présente le potentiel $V_0 + v$ alors que l'électrode opposée présente le potentiel $V_0 - v$.

La force de rappel étant proportionnelle à la différence $(V_0 + v)^2 - (V_0 - v)^2$ est linéaire en v.

Si l'on considère un petit déplacement x de la bille autour de sa position d'équilibre on peut linéariser et admettre que la résultante des forces qui s'exercent sur cette bille est de la forme

$$F = ax - bv ,$$

a et b étant positifs.

Si le détecteur est linéaire, pour un déplacement x il fournit une tension de déséquilibre

$$kx .$$

Si l'on appelle G le gain de la chaîne électronique, on peut écrire

$$v = kGx ,$$

d'où

$$F = (a - bkG)x . \tag{2}$$

Il en résulte qu'en régime permanent l'accélération Γ_x est mesurée par

$$\frac{bkG - a}{m} x . \tag{3}$$

Si l'on s'arrange pour que a soit négligeable devant bkG, pour changer l'étendue de mesure de l'accéléromètre, il suffit d'agir sur le facteur G. C'est le principe que est montré sur la Figure 4.

3. ETAT ACTUEL DE L'ETUDE SUR L'ACCELEROMETRE ONERA

La relation (1) montre que, pour une pulsation propre donnée, l'appareil est d'autant plus sensible que l'on peut mesurer un déplacement x plus petit. Un calcul élémentaire montre qu'avec un temps de réponse de 1,5 seconde environ, en se plaçant dans les conditions de l'amortissement critique, un seuil de mesure en accélération de 10^{-8}m/s^2 correspond à un seuil de détection en déplacement de bille de 10^{-3} micron. L'obtention d'un tel seuil suppose qu'il a été possible de surmonter les difficultés qui apparaissent

- dans la réalisation et le montage mécanique
- dans la conception et la mise au point des chaînes électroniques
- dans les essais.

3.1 Partie Mécanique

La nécessité de mesurer les trois composantes du vecteur accélération et le seuil de sensibilité recherché conduisent à imposer des tolérances d'usinage et de montage extrêmement petites.

Par suite du principe capacitif utilisé c'est la valeur moyenne des défauts de surface qui intervient dans le fonctionnement. Cette valeur moyenne est très inférieure à la valeur des défauts localisés. D'autre part, le déplacement de la bille est petit devant la distance interélectrodes et M. Gay a montré[2, 3] que, dans ces conditions, le gain de la chaîne est une facteur de minimisation des influences des défauts mécaniques.

Le rapport entre l'amplitude des défauts mécaniques acceptables et le seuil de détection en déplacement est lié au gain de la chaîne.

Un rapport de 10^3 semble pouvoir être accepté sans que les performances visées soient remises en cause.

Cependant l'étalonnage de chacune des composantes peut se trouver faussé par une rotation de bille si celle-ci n'est pas une sphère parfaite. L'erreur qui en résulterait serait d'autant plus faible que le défaut de sphéricité de la bille serait plus faible. Il y a donc intérêt à utiliser le procédé d'usinage susceptible de produire les billes présentant le défaut de sphéricité minimum. Jusqu'à maintenant, les meilleurs résultats ont été obtenus par un procédé de rodage dérivé de celui qui est utilisé pour fabriquer des lentilles ou des miroirs optiques. Les cages sont obtenues par le même procédé de rodage. Dans ces conditions les défauts de sphéricité observés sont de l'ordre de 10^{-1} micron.

La précision d'usinage et de montage est obtenue par l'emploi de procédés de contrôle présentant une sensibilité équivalente à celle qui est recherchée pour l'accéléromètre.

Ces procédés sont fondés sur l'utilisation de montages mécaniques spéciaux associés à des mesures capacitives.

Pour éviter les erreurs qui pourraient résulter d'une variation de température au cours du montage ou du contrôle dimensionnel, ces opérations sont effectuées dans une salle à température constante.

Dans ces conditions, il est possible d'obtenir une précision globale de 1 micron.

Les caractéristiques dimensionnelles sont indiquées en microns sur la Figure 5 où un seul axe de sensibilité est représenté.

Plus la distance entre la bille et les électrodes de détection est petite, plus la sensibilité de détection des deplacements de la bille est grande. Au point de vue de la sensibilité il y a donc intérêt à choisir une distance bille-électrode de détection aussi petite que possible. Mais plus cette distance est petite, plus le fonctionnement du dispositif est influencé par les défauts mécaniques d'usinage ou de montage. Il y a donc lieu de choisir un compromis et l'expérience a montré qu'une distance de 30 microns environ représentait une valeur acceptable.

Si l'accéléromètre a un seuil de détection en déplacement de 10^{-3} micron et s'il a un domaine de fonctionnement égal à 10^3 fois le seuil de détection, le déplacement de la bille est compris entre 10^{-3} micron et 1 micron. Il serait donc souhaitable de limiter la course libre de la bille à une valeur légèrement supérieure à 1 micron.

En effet, plus la course libre est grande plus il faut de temps pour amener la bille de sa position de butée à sa position de mesure et par conséquent plus il faut de temps pour mettre l'accéléromètre en condition de mesure.

La course libre de la bille est donc réduite par un jeu de butées.

En pratique il n'est pas possible de réduire cette course à une valeur voisine de 1 micron en raison des tolérances d'usinage et sa valeur est réglée à 20 microns environ.

La position des électrodes d'action est déterminée par la précision recherchée pour l'accéléromètre.

En effet, supposons que l'accélération à mesurer soit nulle et considérons un des axes de sensibilité XX' par exemple (voir Figure 6).

Soient D_1D_2 les positions des électrodes de détection et A_1, A_2 les positions des électrodes d'action.

Soit 0_D la position du centre de la bille qui donne un signal de détection nul. Supposons que la chaîne d'asservissement ne soit pas fermée mais que les tensions de polarisation V_0 soient appliquées sur les électrodes d'action A_1A_2. L'équilibre des forces d'attraction exercées sur la bille par les électrodes A_1 et A_2 se produit lorsque le centre de la bille est en 0_A. Les imperfections de l'usinage et du montage font que 0_A ne coïncide pas avec 0_D. Supposons que l'écart x_1 entre 0_A et 0_D soit celui de la Figure 6; dans ces conditions, en absence d'accélération, avec la chaîne d'asservissement fermée, le centre de la bille occupe une position d'équilibre 0_V telle qu'il existe un écart x_2 entre 0_V et 0_D. Soit d la distance entre les électrodes d'action et la bille lorsque le centre de celle-ci est en 0_A. L'écart est tel que la tension $V_0 + v$ appliquée en A_2 à la distance $d + x_1 + x_2$ produit sur la bille la même attraction que la tension $V_0 - v$ appliquée en A_1 à la distance $d - x_1 - x_2$.

Puisqu'en 0_V la force résultante appliquée sur la bille est nulle, on tire de (2)

$$a(x_1 + x_2) = bGkx_2 .$$

Si a est negligeable devant bGk,

$$x_2 = \frac{ax_1}{bGk} .$$

a et b sont des coefficients fonction de la géométrie des électrodes. En assimilant celles-ci à des électrodes planes il est facile de montrer que

$$\frac{a}{b} = \frac{V_0}{d} ,$$

d'où

$$x_2 = \frac{V_0 x_1}{dGk} .$$

Pour l'accélération la plus grande à mesurer, la tension v appliquée sur les électrodes doit être au plus égale à V_0. Si l'on admet cette valeur limite et si l'on appelle x_M le déplacement de la bille correspondant on peut écrire

$$Gkx_M = V_0$$

et

$$x_2 = \frac{x_1}{d} d_M .$$

Si l'on admet un seuil de détection de 10^{-3} micron et une étendue de mesure égale à 10^3 fois le seuil on a

$$x_M = 1 \text{ micron} .$$

Si l'on veut éviter les termes compensateurs il faut que x_2 soit situé dans le seuil, donc il faut

$$d \simeq 10^3 \text{ microns} .$$

3.2 Partie Electronique

Les performances rendues possibles par la réalisation mécanique ne doivent pas être détériorées par les chaînes électroniques de détection ou d'action.

3.2.1 Chaîne Electronique de Détection

Avec la géométrie adoptée pour l'accéléromètre, la capacité entre la bille et chaque électrode de détection est voisine de 10 pf. Pour ne pas perdre en performance il faut que la chaîne électronique détecte des variations relatives de cette capacité d'un ordre de grandeur égal a 10^{-5}, soit en valeur absolue des variations de 10^{-4} pf. Dans ces conditions, toutes les précautions doivent être prises pour éviter les variations parasites de capacité. Cela conduit à utiliser des connexions très courtes et rigides entre les électrodes et les circuits d'entrée. En conséquence, la chaîne électronique de détection est montée dans un boîtier relativement massif fixé sur la partie mécanique de l'accéléromètre (voir Figure 7).

Après préamplification la tension de déséquilibre du pont fait l'objet d'une détection synchrone. L'ensemble du schéma de principe de la chaîne électronique de détection est montré sur la Figure 8. Cette chaîne comporte un certain nombre de circuits intégrés qui entraînent une diminution notable du volume. Il faudrait en effet 250 transistors si les chaînes de détection étaient constitués d'éléments séparés.

On peut voir sur l'enregistrement représenté sur la Figure 9 que la stabilité de ces chaînes est très bonne et ne limite aucunement les performances de l'accéléromètre.

3.2.2 Chaîne d'Action

En raison de la distance de 10^3 microns qui sépare la bille des électrodes d'action les tensions appliquées sur ces électrodes sont assez élevées. La tension de polarisation est fixée à 800 V. En limite supérieure d'étendue de mesure la tension $V_0 + v$ atteint 1500 V.

Dans ces conditions toute poussière existant en surface peut être l'amorce d'une décharge disruptive entre électrodes et bille; c'est pourquoi il est impératif d'effectuer le montage de l'accéléromètre dans une très bonne salle dépoussiérée.

Par ailleurs, il faut éviter que la bille prenne une charge électrostatique qui produirait des efforts d'attraction parasites. Pour cela, les électrodes d'action sont doublées et alimentées comme il est montré sur la Figure 10 pour un axe de sensibilité. Cette disposition donne un potentiel résultant nul alors que les efforts produits par chaque paire d'électrodes s'ajoutent.

L'amplification est effectuée sur une fréquence porteuse de 40 kHz. Un transformateur de sortie accordé alimente le montage redresseur qui fournit les tensions appliquées sur les électrodes. La stabilité du gain est obtenue par contre-réaction. Le schéma de principe fonctionnel d'une chaîne électronique d'action est représenté sur la Figure 11.

3.2.3 Dispositif de Changement Automatique des Etendues de Mesure

Le changement d'étendue de mesure de l'accéléromètre s'effectue par changement du gain de la chaîne d'asservissement électronique. Trois étendues de mesure présentant un recouvrement ont été prévues:

- (gamme 1) 10^{-5} m/s^2 à 10^{-2} m/s^2
- (gamme 2) 3.10^{-7} m/s^2 à 3.10^{-4} m/s^2
- (gamme 3) 10^{-8} m/s^2 à 10^{-5} m/s^2.

Pour certaines applications spatiales il peut être très difficile de prévoir et de mettre en oeuvre un programme de commutation des gammes de sensibilité; aussi il a été décidé de doter l'accéléromètre d'un dispositif de commutation automatique.

Le principe d'un tel dispositif est simple: il suffit de définir une limite inférieure et une limite supérieure de détection. Chaque fois que la grandeur détectée est extérieure à ces limites le dispositif commande un changement d'étendue de mesure dans le sens convenable. En raison des couplages existant entre les trois axes de sensibilité il faut que les étendues de mesure soient les mêmes sur X, Y et Z. Il est donc nécessaire d'effectuer les commandes de changement d'étendue de mesure avec des circuits logiques: circuits "et" pour augmenter la sensibilité, circuits "ou" pour la diminuer. Le schéma fonctionnel de l'ensemble du dispositif est représenté sur la Figure 12. Ce dispositif est actuellement parfaitement au point. La Figure 13 montre un exemple d'enregistrement obtenu lors de deux changements de gamme consécutifs. La trace supérieure montre le mouvement réel de la bille partant d'une position de butée pour venir au centre de l'étendue de mesure. La trace inférieure montre, par l'emplacement de tops sur le signal de mesure, les deux changements de gamme.

3.3 Essais de Laboratoire

Ainsi qu'il a été expliqué précédemment[4], le couplage entre les axes de sensibilité X, Y, Z ne permet pas d'effectuer les essais de l'accéléromètre dans les conditions d'environnement de la pesanteur.

La technique d'essais en chute libre dans le tube d'impesanteur de l'ONERA[5] est maintenant parfaitement opérationnelle.

Grâce à cette technique il a été possible d'effectuer de nombreux essais et de procéder à une mise au point progressive de l'accéléromètre. La Figure 14 donne un exemple des courbes de décélération obtenues sur les trois axes de sensibilité lors d'une chute libre dans le tube.

Les courbes permettent d'analyser le fonctionnement de l'accéléromètre et éventuellement de calculer un étalonnage.

Ce calcul s'effectue en comparant la décélération mesurée par l'accéléromètre à celle qui est déterminée en partant de la caractéristique de la vitesse de la capsule mesurée en fonction du temps par un procédé d'interférométrie Laser[6].

Cette méthode d'étalonnage ne donne pas entière satisfaction pour deux raisons:

- la précision finale n'est pas suffisante
- la capsule d'essais en chute libre n'est pas équipée pour permettre d'effectuer simultanément un essai d'accéléromètre et une mesure de vitesse.

Aussi la plupart des étalonnages effectués jusqu'à maintenant l'ont été en procédant de la manière suivante:

- un effort de retenue est exercé sur la capsule en cours de chute par un fil élastique en caoutchouc,
- cet effort F est mesuré par un dynamomètre piézoélectrique et la mesure est enregistrée en fonction du temps. Connaissant la masse totale M de la capsule on peut déterminer en fonction du temps la décélération $\Gamma = F/M$ subie par la capsule du fait de la traction exercée par le fil. Le fil est choisi pour que Γ soit situé à l'intérieur de l'étendue de mesure de l'accéléromètre correspondant à la sensibilité la moins grande,
- la pression à l'intérieur du tube d'impesanteur est réduite à sa valeur minimum afin que le freinage de la capsule par l'air résiduel soit négligeable devant le freinage imposé par le fil.

A l'avenir il est envisagé d'effectuer les étalonnages physiques d'une manière plus directe. Le principe retenu est fondé sur la mise en rotation à l'intérieur de la capsule d'essai en chute libre, d'une masse dont les caractéristiques mécaniques sont connues avec précision et dont le mouvement est enregistré.

Lorsque ce nouveau procédé d'étalonnage sera au point, il restera à qualifier l'accéléromètre pour un emploi spatial. Nous pensons que cette qualification ne peut être décidée qu'à la suite d'un essai en vol. Un tel essai a été défini et sa préparation est commencée.

4. DEFINITION DE L'ESSAI DE QUALIFICATION EN VOL

Cet essai, dont les modalités sont actuellement en cours de discussion avec le CNES (Centre National d'Etudes Spatiales), prévoit le lancement d'une capsule sphérique contenant l'accéléromètre, par fusée sonde, sur une trajectoire culminant à 200 km d'altitude environ.

La forme sphérique de la capsule permettra de corréler les indications de l'accéléromètre avec la valeur de la densité atmosphérique. Pour que l'expérience soit facile à dépouiller il faut que les décélérations dues au mouvement de la capsule soient inférieures au seuil de mesure. Un calcul montre que cela est possible en prenant les précautions suivantes:

- montage de l'accéléromètre au centre de gravité de la capsule avec une précision égale à $\pm 10^{-1}$ mm.
- réduction du module de la vitesse angulaire de la capsule à 10^{-2} rad/s par emploi de deux gyroscopes agissant par effet mécanique direct.

Dans ces conditions, une trajectoire théorique, comme celle qui est représentée sur la Figure 15, permet de calculer à partir de la constitution de l'atmosphère standard, la décélération qui serait imposée à une capsule de dimensions et de masse données. Les résultats d'un tel calcul sont représentés en fonction du temps sur la Figure 16.

Les trois étendues de mesure permettront de bénéficier d'une durée de mesure satisfaisante.

Au cours de l'essai en vol les changements de gamme seront mis à profit pour envoyer la bille sur les butées afin d'arrêter son éventuel mouvement de rotation par rapport à la cage et pour effectuer un étalonnage électrique par introduction dans la chaîne de mesure d'un signal d'amplitude donnée. La Figure 17 montre l'enregistrement obtenu lors d'une simulation de ces différentes opérations.

REFERENCES

1. Hennion, P. *La Mesure des Très Faibles Accélérations.* ONERA Note Technique No.89, 1965.

2. Gay, M. *Recherche sur un Principe d'Accéléromètre de Grande Sensibilité.* La Recherche Aérospatiale, No.110, 1966.

3. Gay, M. *Accéléromètre à Grande Sensibilité.* Communication présentée au XVIIème Congrès International d'Astronautique, Madrid, 9-15 octobre 1966.

4. Delattre, M. *L'Accéléromètre ONERA à Grande Sensibilité.* Congrès Mesucora, Paris, 17-21 avril 1967. TP ONERA 449.

5. Delattre, M.; Dubois, G. *Mise en Place à l'ONERA d'une Installation d'Essais en Impesanteur.* La Recherche Aérospatiale, No.110, 1966.

6. Taran, J.P. *Mesure Précise des Vitesses au Moyen d'un Interféromètre à Laser.* La Recherche Aérospatiale, No.119, 1967.

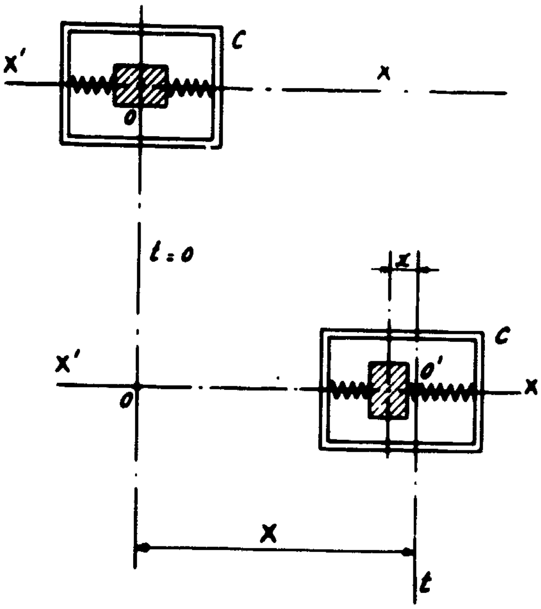


Figure 1

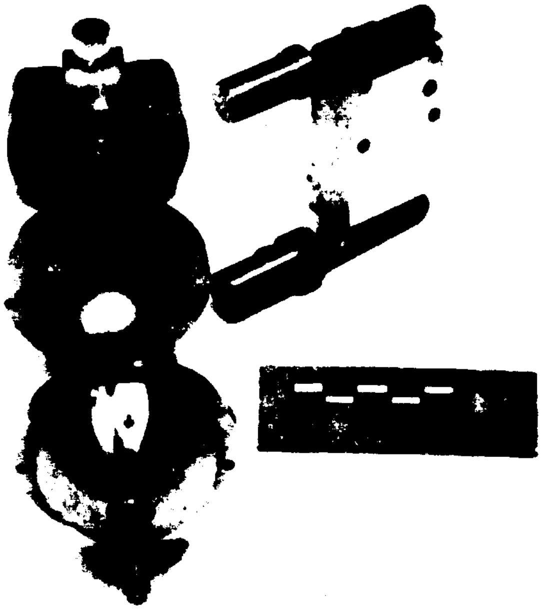


Figure 2

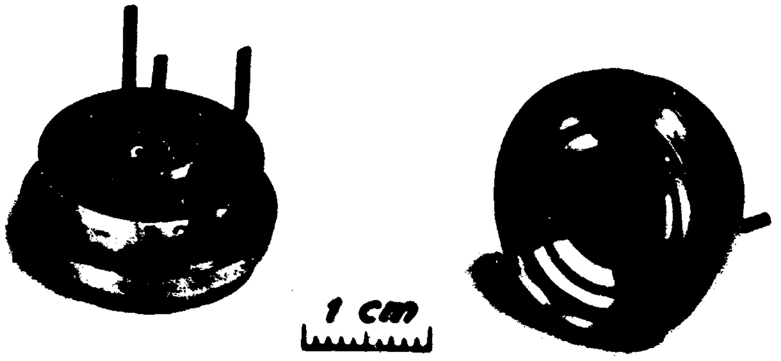


Figure 3

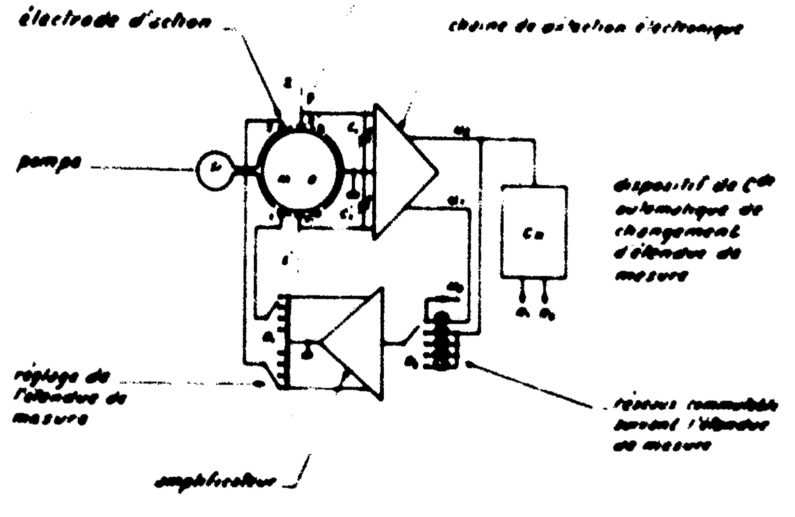

électrode de détection
électrode d'action
chaîne de détection électronique
pompe
Ca
dispositif de Cde automatique de changement d'étendue de mesure
réglage de l'étendue de mesure
réseaux commutables suivant l'étendue de mesure
amplificateur

Figure 4

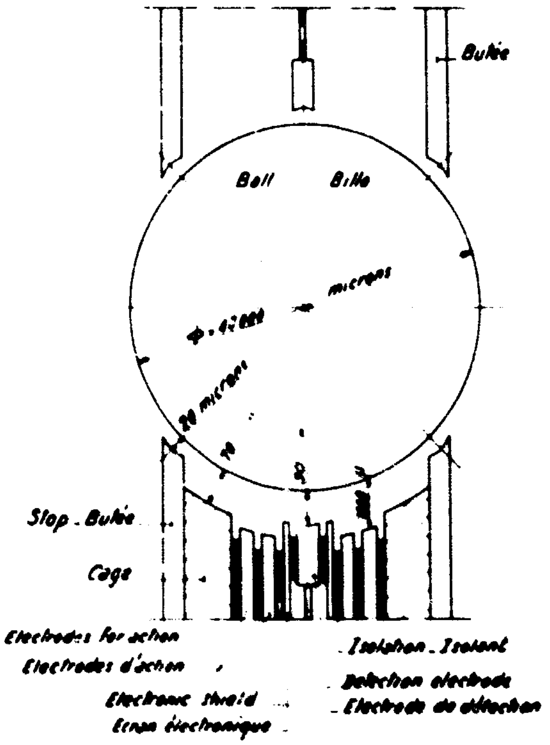


Figure 5

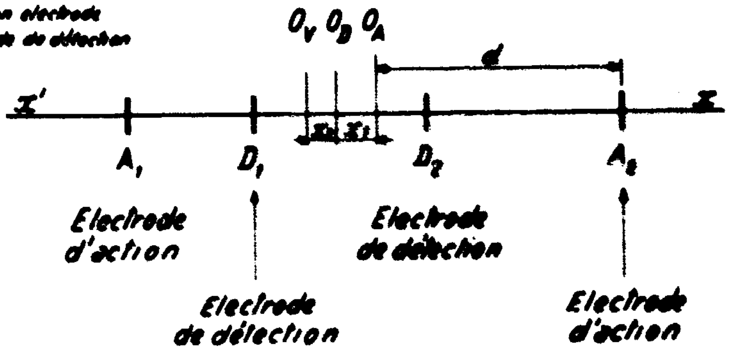


Figure 6

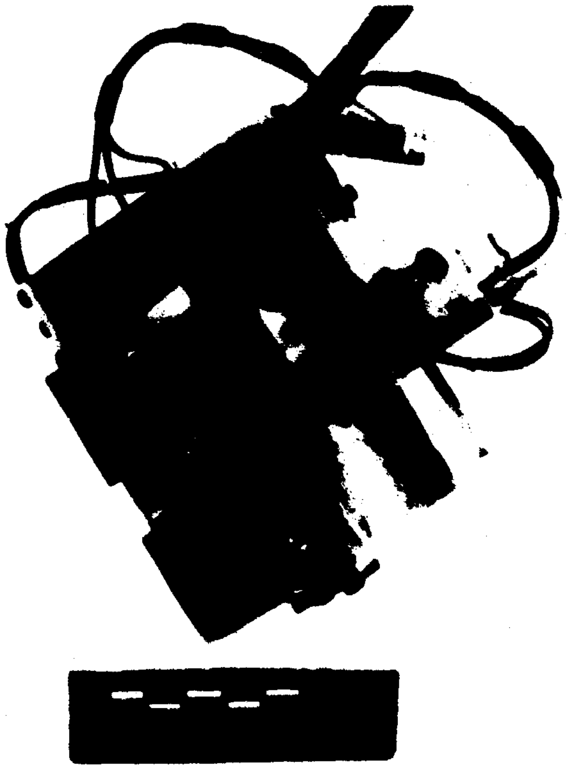


Figure 7

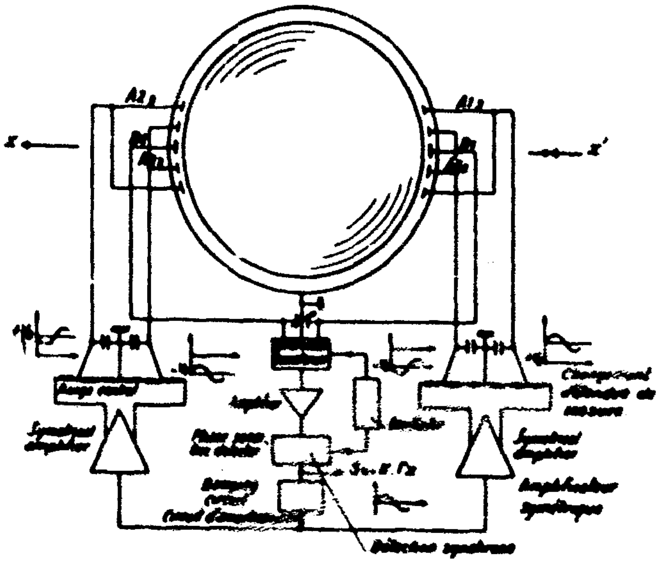


Figure 8

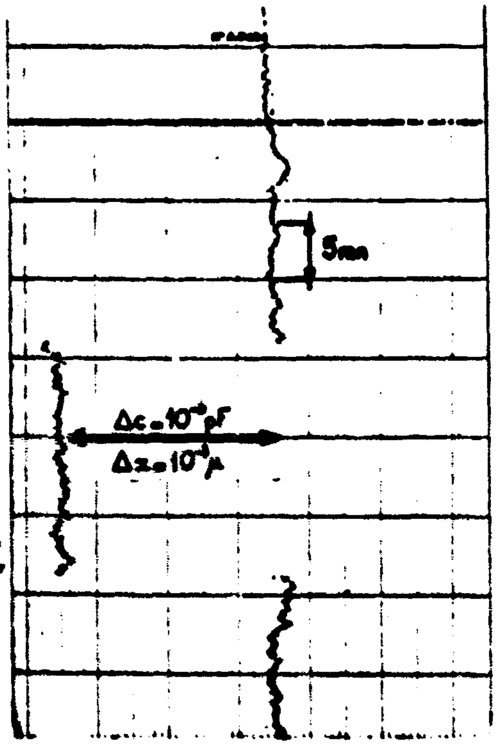


Figure 9

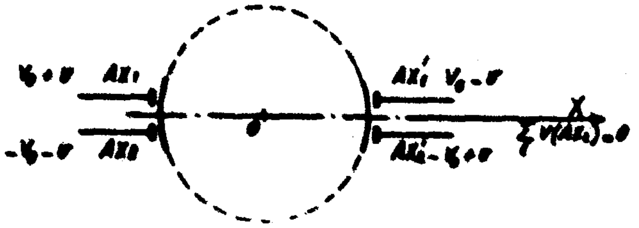


Figure 10

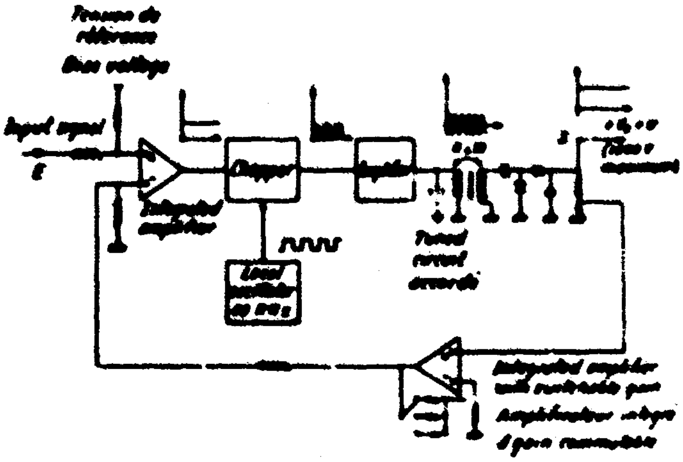


Figure 11

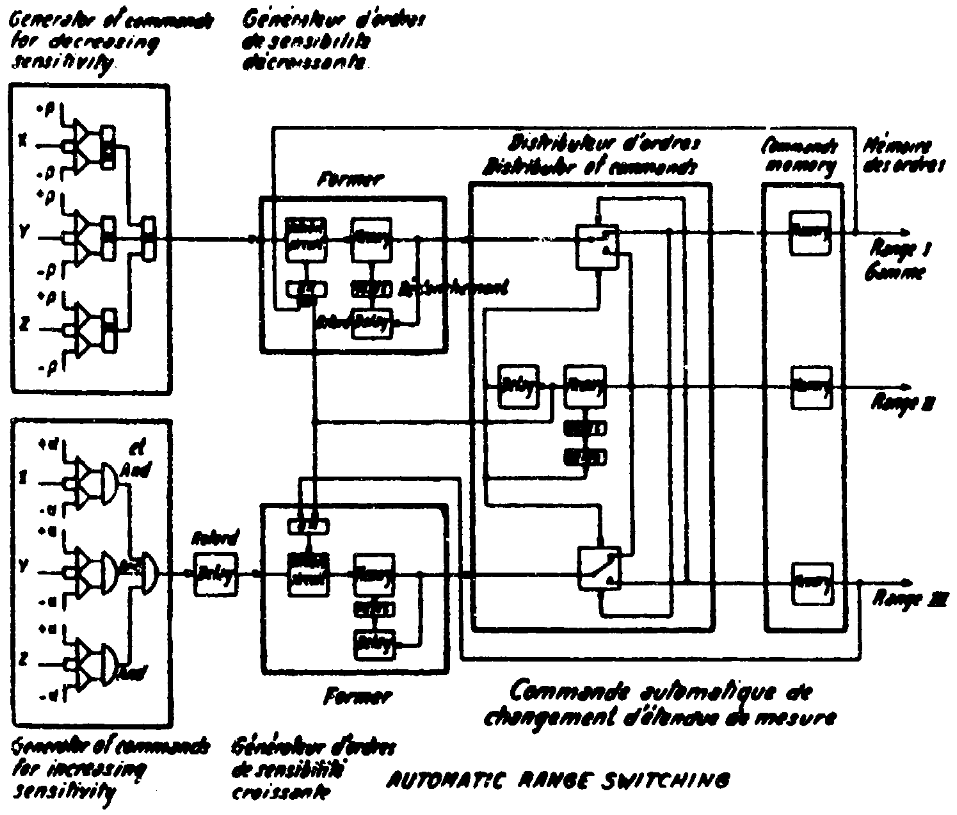

Generator of commands for decreasing sensitivity
Générateur d'ordres de sensibilité décroissante
X
Y
Z
+p
-p
Former
Distributeur d'ordres
Distributor of commands
Commands memory
Mémoire des ordres
Memory
Delay
Range I
Gamme
Range II
Range III
+a
-a
et
And
Retard
Delay
Former
Commande automatique de changement d'étendue de mesure
Generator of commands for increasing sensitivity
Générateur d'ordres de sensibilité croissante
AUTOMATIC RANGE SWITCHING

Figure 12

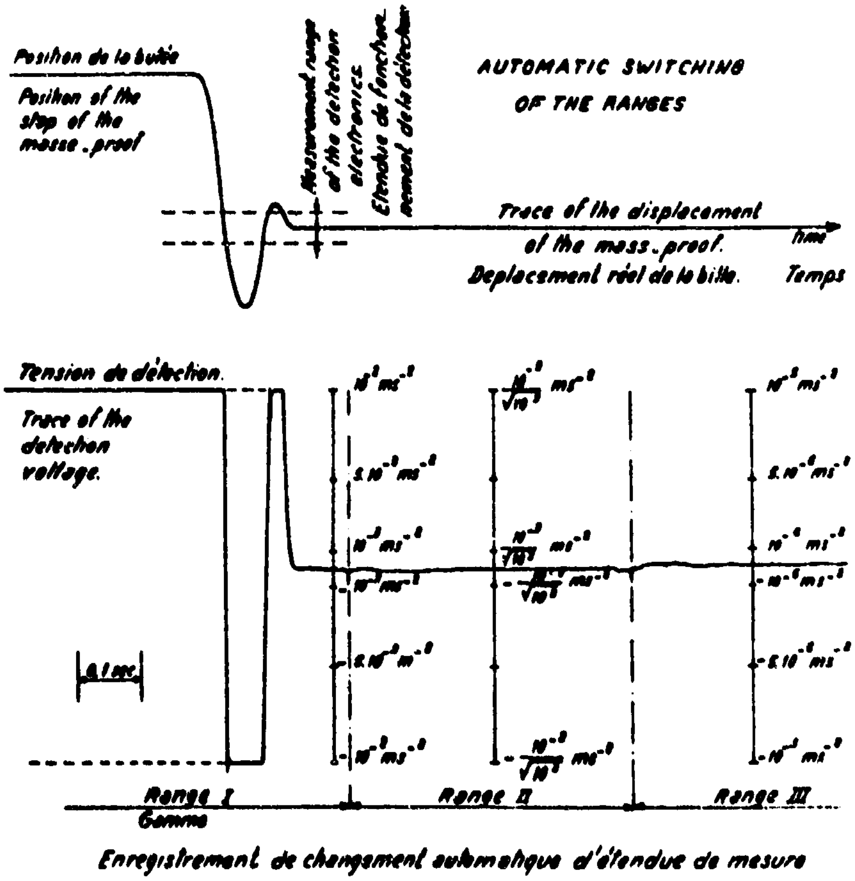


Figure 13

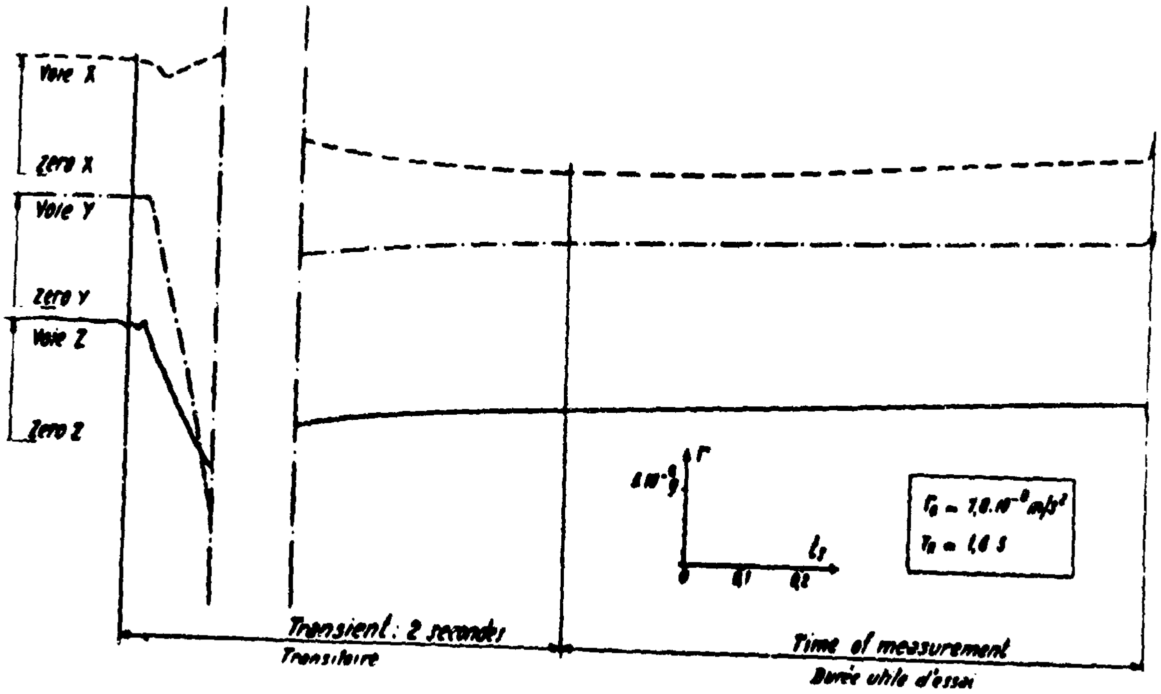


Figure 14

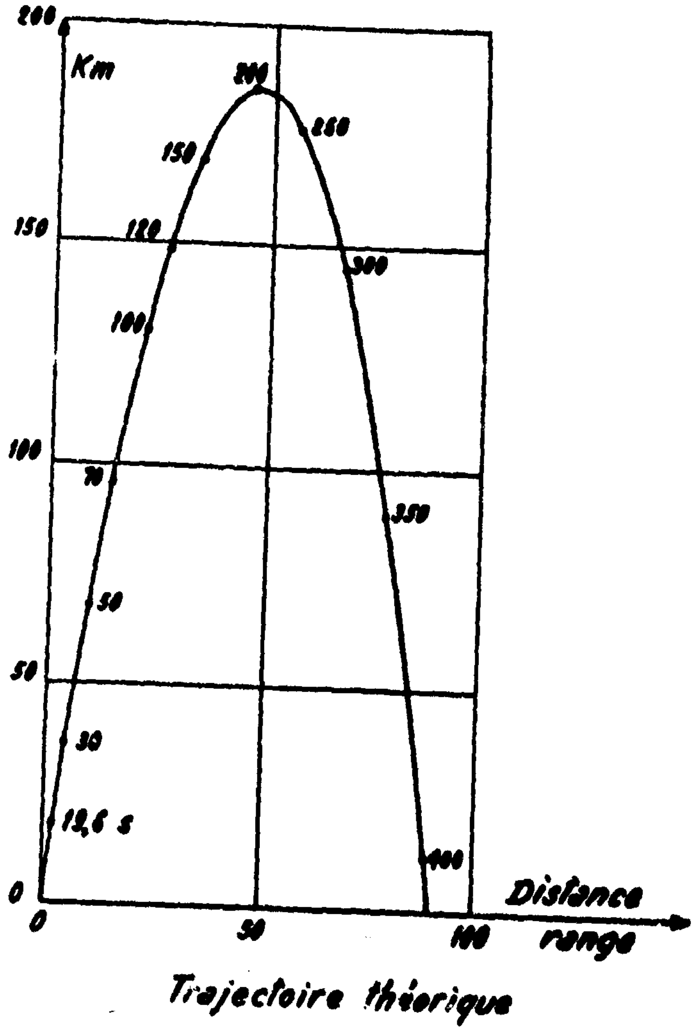
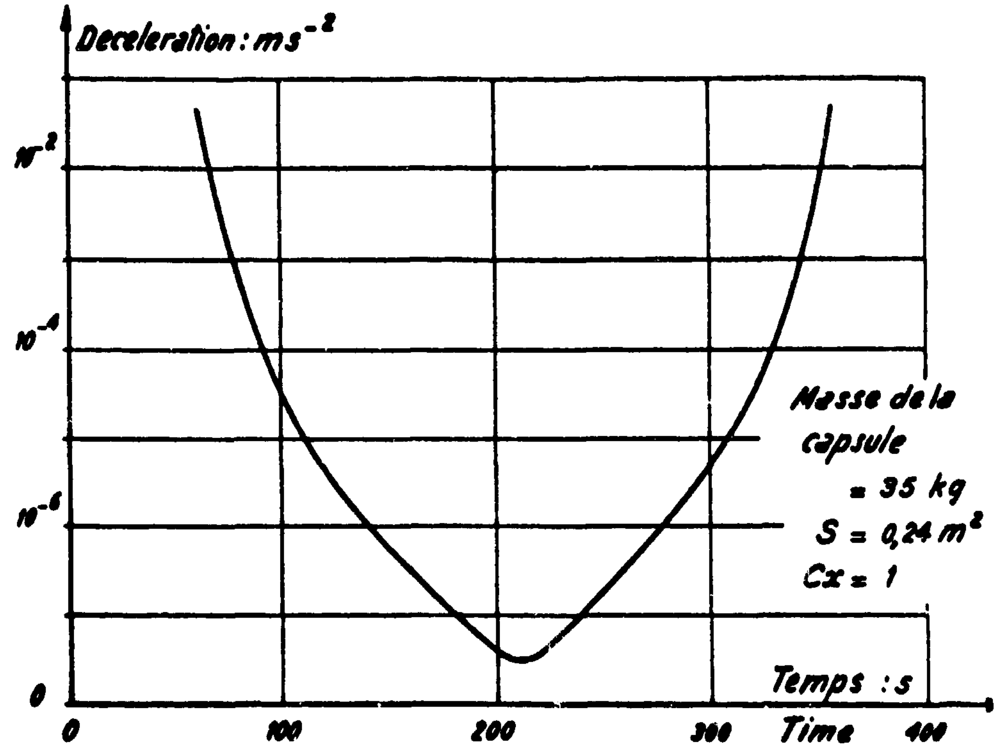


Figure 15



Essai en vol

Figure 16

SIMULATION OF THE CALIBRATION IN FLIGHT
Simulation d'un étalonnage en cours de l'essai en vol

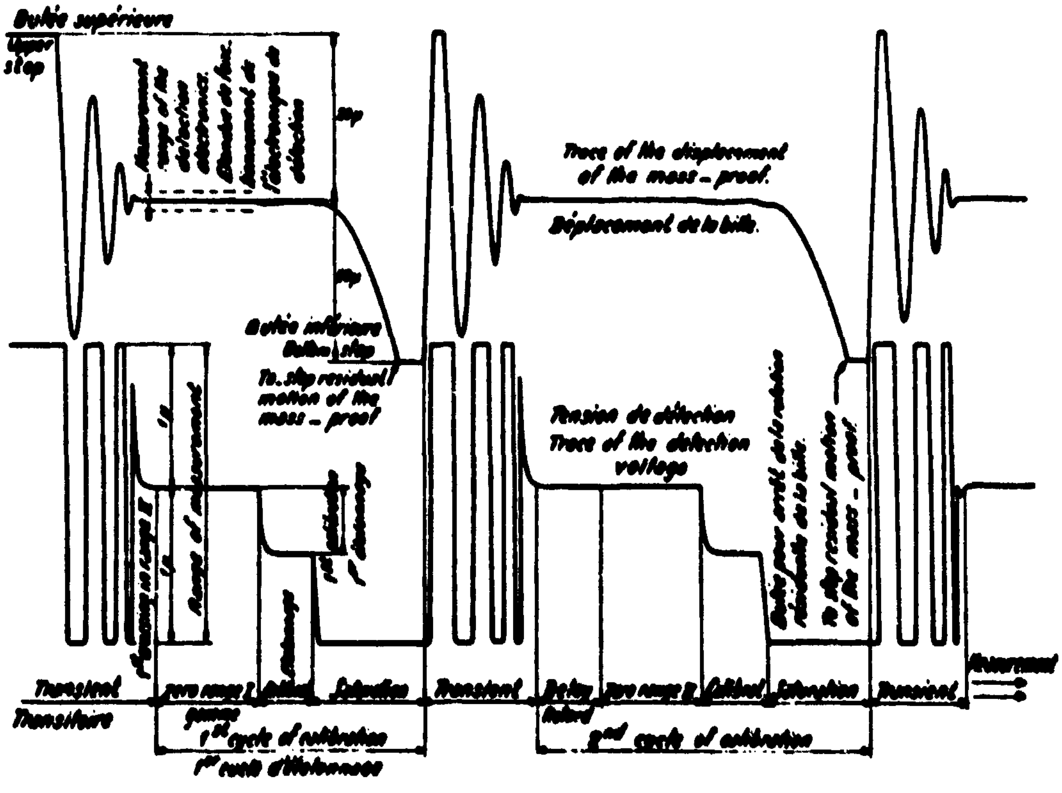


Figure 17

THE VIBRATING STRING ACCELEROMETER

by

Robert O. Bock

American Bosch Arma Corporation
Santa Ana, California 92705, USA

THE VIBRATING STRING ACCELEROMETER

Robert O. Bock

To a physicist, the simple word "string" means a piece of material of appreciable length but small cross-section; it possesses tensile strength properties, but negligible shear or stiffness characteristics. A fine watch chain might be a good example.

Such a physical string, clamped rigidly at the ends of segment length L and subjected to tension T, will exhibit a natural frequency of transverse vibration, i.e. displacement in the plane normal to the string length, given by Rayleigh and others by the expression

$$f = \frac{1}{2L}\sqrt{\left(\frac{T}{m/L}\right)} , \tag{1}$$

where m is the mass of the string. The expression is readily derived from a simple differential equation used as a standard exercise for the beginning student in Theoretical Mechanics. (Harmonic modes of vibration are also possible but are of only passing interest at this time.)

Accelerometers may be thought of as devices which transduce the quantity Force resulting from Newton's Second Law of Motion,

$$F_x = M\frac{d^2x}{dt^2} , \tag{2}$$

into some convenient analog, usually electrical. It is thus apparent that the vibrating string not only has potential as the desired conversion device but produces a resulting analog, i.e., frequency, which is the most accurately measurable of all physical quantities.

We proceed to examine practical configurations and inherent qualities of such Vibrating String Accelerometers.

Consider first a simple mass hanging on a wire, which wire will be caused to vibrate in the transverse mode previously described. Assume a fine cross-support system and preload tensioning device for convenience; Figure 1 illustrates such a system. The string will be maintained in a state of low-amplitude continuous vibration by having it in a strong permanent magnet field and by passing a small, limited electric current through it by means of a connected feed-back amplifier. The amplifier functions much as a regenerative radio receiver feeding back in-phase power to its input antenna. (More detailed amplifier requirements will be examined later.)

From Equation (1) it can be seen that the string frequency in the absence of any accelerations (or gravitational effects) to the system is given by

$$f_1 = \frac{1}{2L}\sqrt{\left(\frac{T_0}{m/L}\right)} . \tag{3}$$

where T_0 = "preload tension".

If now the system is physically accelerated to the right, we may write

$$f_1 = \frac{1}{2L}\sqrt{\left(\frac{T_0 + Ma}{m/L}\right)} , \tag{4}$$

where M = mass attached to the string

a = acceleration,

or, using a series expansion,

$$f_1 = C_0 + C_1 a + C_2 a^2 + C_3 a^3 + \ldots . \tag{5}$$

The C's are constants containing m, L, T_0, M.

The feasibility of utilizing this device depends on the rate of convergence of this series, since non-linearity can be a handicap for many uses.

The ratio C_2/C_1, for example, is given by

$$\frac{C_2}{C_1} = \frac{1}{4}\frac{M}{T_0}$$

and sample constants for an instrument* might be taken as (assume rectangular cross-section)

string length = 2 cm

string width = 0.01 cm

string thickness = 0.002 cm

string density = 10 gm/cm^3

mass = 10 gm

tension = 600 gm.

These values yield constants approximately as follows:

C_0 = 13,650 Hz

C_1 = 114 Hz/g

C_2 = 0.47 Hz/g^2

C_3 = 0.004 Hz/g^3.

Convergence of this series is obviously not rapid enough for applications such as missile guidance, where high precision is required and acceleration values exceed 1 g by appreciable amounts. However, for very small dynamic range requirements, e.g., gravity measurement, this non-linearity poses no problem. Instruments of the type described have been used very successfully for gravity measurement on submarines and on land; literature describing several of these applications is given in the References at the end of this paper.

To improve linearity of the output of the accelerometer described, one may decrease M, increase T_0, or apply compensation or calibration techniques. In general, a more satisfactory configuration than that shown is far more successful.

Let us add to the instrument described an identical "mirror image" attached at the left end of the coil spring. The result is shown in Figure 2, with some further construction details shown in Figure 3. To develop the equation for this accelerometer, we first note that an acceleration that is positive for the original string (on the right) has a negative sense for the second string (on the left); Equation (4) then becomes, for f_2,

---

* These values do not represent those of any actually known strings or instruments built.

$$f_2 = \frac{1}{2L_2}\sqrt{\left(\frac{T_0 - M_2 a}{m_2/L_2}\right)} , \qquad (6)$$

and the counterpart of Equation (5) becomes

$$f_2 = C_{0_2} - C_{12}a + C_{22}a^2 - C_{32}a^3 + \dots . \qquad (7)$$

where the "2" subscripts refer to the second string. If we now *subtract* the two string frequencies, we have the new output equation for this two-string accelerometer:

$$f_1 - f_2 = (C_{0_1} - C_{0_2}) + (C_{11} + C_{12})a + (C_{21} - C_{22})a^2 + (C_{31} + C_{33})a^3 + \dots ,$$

or, more simply,

$$f_1 - f_2 = K_0 + K_1 a + K_2 a^2 + K_3 a^3 + \dots . \qquad (8)$$

We note the following for this "difference frequency" equation:

(i) Even order terms are zero within our ability to match the two halves, i.e., to make $C_{n1} = C_{n2}$.

(ii) The third-order term cannot be made equal to zero, since single string coefficients of the same sign are *added* to obtain the final coefficient.

Production experience has demonstrated that K_2 values are easily held to a few parts per million, i.e., a few micro g per g^2. K_0 is of little consequence so long as it remains fixed, in fact it is usually set arbitrarily at a selected non-zero value for convenience in use.

The third-order term is usually in the 10 to 40 micro g per g^3 region, depending on the application requirements and the resulting choice of parameters for instrument design.

Using the constants of the single string hypothetical instrument previously discussed, we obtain the following (approximately). For example:

K_0 = 0 to 2 Hz (say)

K_1 = 230 Hz/g

K_2 = -

K_3 = 35 micro g/g^3.

To characterize this instrument, we make use of three terms which we define here as follows:

1. *Sensitivity* is the smallest input signal to which the instrument will exhibit proper response.
2. *Accuracy* is the precision with which a calibration can be usefully repeated.
3. *Resolution* is given by the smallest increment of acceleration (or, as we shall see, velocity) which can be usefully read out.

For the first of these quantities, the theoretical sensitivity falls in the region of thermo-molecular noise level of the mechanical parts; in practice, the degree of sophistication of readout equipment has limited useful sensitivity to date in the 10^{-8} g region.

The accuracy achieved for any instrument depends heavily on the dynamic range expected of it; for the 20 g range, accuracy of the order of 1 part per million has been accomplished, while for the 0.01 to 0.1 g dynamic range accuracy has approached 1 part in 100 million.

Resolution is a matter of readout technique, and a variety of electronic "counting" schemes have been employed. Note first that the output of the accelerometer is true frequency modulation and originates on two "carrier" frequencies, i.e., f_1 and f_2. (It is also worth noting that several other unique features exist for this instrument:

1. There is no servo system.

2. Total power dissipated in the sensor is less than 1 micro-watt and this is invariant, regardless of input signal.

3. There are no moving parts in the usual sense - string amplitudes are usually in the 100 microinch region.

4. Integration to obtain velocity (often more significant than acceleration) is accomplished without error with no converters, integrators, or added devices beyond the standard micro-sized electronic "flip-flop" counters.

5. No difficult conversions between electrical and mechanical quantities occur at all.)

Three readout systems in use will be described; all relate significantly to resolution. Other methods or varying combinations of these methods may also be feasible.

Figure 4 shows the first and probably simplest readout method. This method of counting individual string vibrations and digitally subtracting is most feasible in general where the accelerometer is in a system utilizing a digital computer, which supplies the built-in computation ability. It requires that the inherent accelerometer resolution be adequate for the application without addition of the techniques shown below. Note that the output may be either velocity or acceleration if the sample time intervals are known.

The above method used in the laboratory has a pitfall for those seeking ultimate precision. If the time intervals for sampling the separate strings are not identical, then each string sensor finds itself at a different phase of the ever-present seismic disturbances in the laboratory and only very long samples can hold this reading error to very small values. The long times are apt to be inconvenient.

A superior laboratory readout system in widespread use is shown in Figure 5. In this system, a simple heterodyne circuit subtracts the two string frequencies and delivers the difference to a divider section in a commercial electronic counter. Every nth wave gates the counter, which registers the number of high frequency clock pulses which pass for an exactly known whole number of difference frequency cycles. This period measurement yields 1 million counts resolution in 1 second; as high as 50 million counts per second have been used. Counting times (note that the divide factor can be chosen at will) can be varied from a fraction of a second to hundreds of seconds - or even days if desired. Precision is limited only by the practical "cleanliness" of counter circuitry and input waves and by the perfection of triggering circuits in the counter. Examples are known of 1500 second counting times, giving resolution of one part in $1500 \times 10 \times 10^6$, or 1 in 1.5×10^{10}.

An interesting variation of this technique has been useful in gravity measurement where the output frequency is nearly constant. Note that the "counter" is actually determining the *ratio* of two frequencies, that from the heterodyne and that of the clock. If the clock frequency is obtained from a frequency "synthesizer", which supplies essentially any desired frequency locked to a precision crystal, then the clock may be set at the heterodyne frequency or, say, 10,000 times the heterodyne value (which is nearly constant), and the total counts accumulated on the register will be unity followed by all zeros. Values may be set so that an observed change of 1 count represents exactly 1 micro g (or 0.1 micro g) as desired.

Below is a tabulation of numbers taken at random from a laboratory counter and printer illustrating the use of the technique described above. A vibrating string instrument used as a gravity meter supplied the results. Counting time was several hundred seconds.

```
1 0 0 0 0 0 0 6
1 0 0 0 0 0 0 6
1 0 0 0 0 0 0 6
1 0 0 0 0 0 0 6
1 0 0 0 0 0 0 6
1 0 0 0 0 0 0 5
```

The third readout system in use is shown schematically in Figure 6. The first portion of the figure shows how the resolution of the instrument is first increased by 100 times by electronic frequency multiplication; in our hypothetical instrument, previously described, the equivalent scale factor would be increased from 230 Hz/g to 23,000 Hz/g. One may choose to heterodyne and multiply further if he wishes; the only limits are dynamic requirements, essentially, for the application. The multiplied frequency, for example, is really derived by extrapolation of the phase properties of each carrier wave cycle through the next period of the carrier. Should the input wave be so drastically modulated during one cycle (in 1/13,650 sec in our string-carrier example), then the phase detection circuit may malfunction In over a decade of continuous use of these circuits, only one situation has been encountered in which detectable errors occurred from such an event: a small oceanographic ship detected several momentary "phase lock" errors one day; the occurrence was quickly correlated with depth-charge-like explosions being set off directly under the ship for seismic sounding experiments.

The second portion of Figure 6 illustrates one of the commonly employed techniques for frequency multiplying. A voltage controlled oscillator is nominally adjusted to put out a signal of frequency $n \times f$, where f is the "input" frequency. A feedback loop divides this output frequency by n digitally, i.e., using "flip-flops", with zero error, and the phase detector circuit matches every n^{th} wave of the output to the input f. Error voltage is used to keep the voltage controlled oscillator locked to the proper frequency in this pure electronic servo.

Auxiliary features of some interest have been incorporated in a number of instruments for special applications. One of these is an automatic zero adjust circuit employed for certain space-flight purposes.

The driving amplifier for the vibrating string has the following requirements:

1. Phase shift between signal and feedback to the string must be kept small.
2. Voltage to the string must be reasonably constant.
3. Limited harmonic content is permitted in the feedback wave.
4. Considerable bandpass must usually be provided to permit proper response to sideband frequencies.

Note the first of these: when extreme stability of zero bias is required for say, some months in space, then the circuit of Figure 7 is used to automatically provide true zero for bias so long as the accelerometer is in free fall. Should the difference frequency not be zero, the phase difference is detected, amplified, integrated and used to phase shift the feedback voltage to one string until perfect frequency match is restored. The phase shifter is a simple electronic circuit. When it is desired to read acceleration, the switch (electronic) leading to the integrator is simply opened.

It would be unfair to omit mention of one possible result of the cubic term non-linearity. If the hypothetical instrument previously described were subjected simultaneously to acceleration which is steady or slowly changing, and to vibration of a frequency higher than the sampling rate of readout, then rectification errors may result. This situation is readily visualized in the sketch of Figure 8 (grossly exaggerated). If readout samples average over several vibration cycles, then an incorrect average is obviously obtained. The exact expression for this (error in K_1) is given by

$$3K_3a_v^2 \, , \tag{9}$$

where a_v is the root mean square vibration amplitude. Thus, for our hypothetical instrument at 1 g steady and 0.1 g 0 to peak vibration, we have

$$\epsilon_{K_1} = 3 \times 35 \times (0.7 \times 0.1)^2 \text{ micro g/g} \, .$$

or 0.5 micro g/g (or 5 parts in 10 million).

While several correction schemes have been used, a particularly interesting one has been developed for the vibrating string gravity meter. If we substitute the quantity $g + c$, where c is deviation from a normal 1 g value (say on a ship), for a in the equation for the accelerometer output and write a new series expansion about the 1 g point instead of the 0 g point (as in Equation (8)), we achieve an equation whose non-linearities are in terms solely of c, the variations from 1 g, i.e.,

$$f_1 - f_2 = K_0' + K_1'c + K_2'c^2 + K_3'c^3 + \ldots .$$

Physically, to satisfy the requirement that $K_2' = 0$, we find that the detailed expressions (in terms of masses and string parameters) for K_2' and K_3' call for a quantitative *mismatch* of the two halves of the instrument, which has not been difficult to achieve in practice.

The whole advantage of this modification lies in the fact that the rectification errors are zero about the assumed 1 g point about which the expansion is derived.

Security restrictions limit the publishing of actual data on vibrating string accelerometers in many cases. Figures 9 and 10, however, show results of some early tests employing an unclassified vibrating string accelerometer used as a sea-borne gravimeter and as a tide meter, respectively. In both cases, off-the-shelf 30 g instruments were used.

In summary, the vibrating string accelerometer may be said to offer high precision performance with great simplicity and reliability. There are numerous special production techniques required to achieve high yield rates of instruments, but these have been solved with no recourse to super accurate tolerances or exotic processes. Technical advances are still being made in numerous features and in utility of these sensing instruments.

REFERENCES

1. Gilbert, R.L.G. *A Dynamic Gravimeter of Novel Design.* Proceedings of the Physical Society, London; Vol.62, Series B, 1949, pp.445-454.

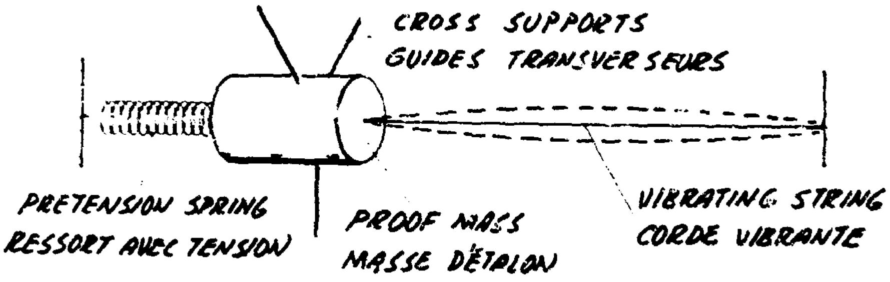
2. Gilbert, R.L.G. *Gravity Observations in a Borehole.* Nature, Vol.170, 1952, pp.424-425.

3. Lozinskaya, A.M. *The String Gravimeter for the Measurement of Gravity at Sea.* Bulletin of the Academy of Sciences, USSR, Geophysical Series.English Translation No.3, 1969 pp.398-409.

4. Tomoda, Y., Kanamori, H. *Tokyo Surface-Ship Gravity Meter α-1. Collected Reprints,* The Ocean Research Institute, University of Tokyo, Vol.1, 1962, pp.116-145.

5. Goodell, R.R., Fay, C.H. *Borehole Gravity Meter and its Application.* Geophysics, Vol.29, 1964, pp.774-782.

6. Wing, C.G. *An Experimental Deep Sea-Bottom Gravimeter.* Journal of Geophysical Research, Vol.72, No.4, 1967, pp.1249-1257.



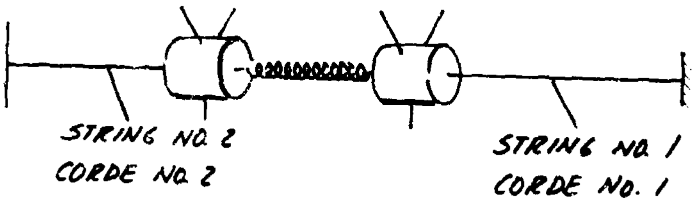
ZERO ACCELERATION :
ACCELERATION NULL:

$$f = \frac{1}{2L}\sqrt{\frac{T_0}{m/L}}$$

POSITIVE ACCELERATION:
ACCELERATION POSIT

$$f = \frac{1}{2L}\sqrt{\frac{T_0 + Ma}{m/L}}$$

Figure 1



$$f_1 = \frac{1}{2L}\sqrt{\frac{T_0 + Ma}{m/L}}$$

$$f_2 = \frac{1}{2L}\sqrt{\frac{T_0 - Ma}{m/L}}$$

Fig. 2 Vibrating string accelerometer
Accelerometeur aux cordes vibrantes

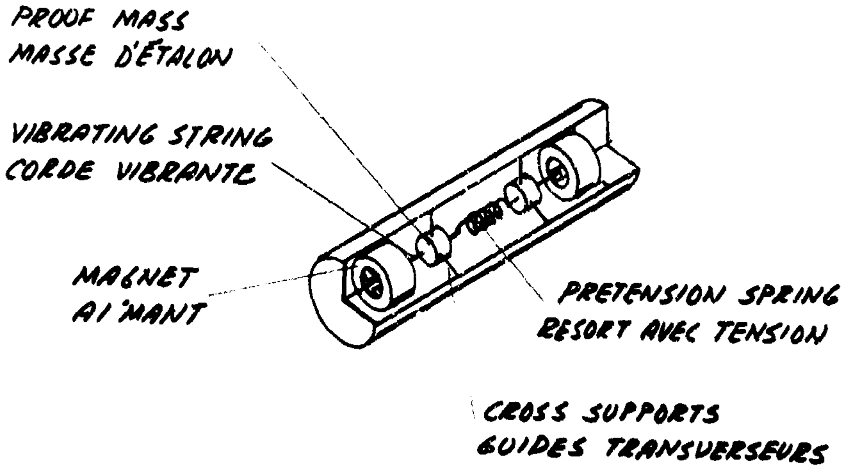


Figure 3

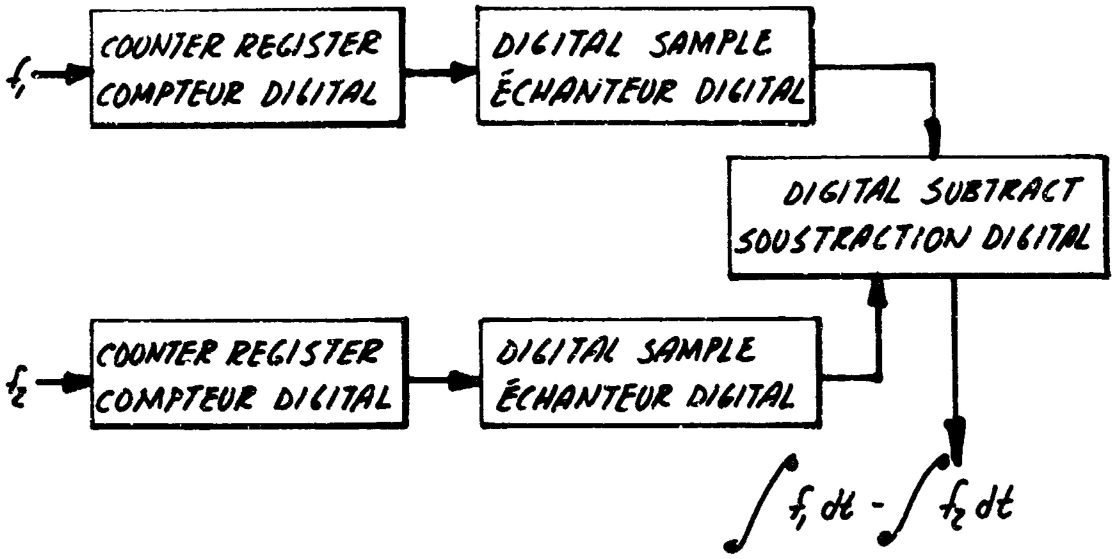


Fig. 4 Readout method No. 1
Méthode de présentation No. 1

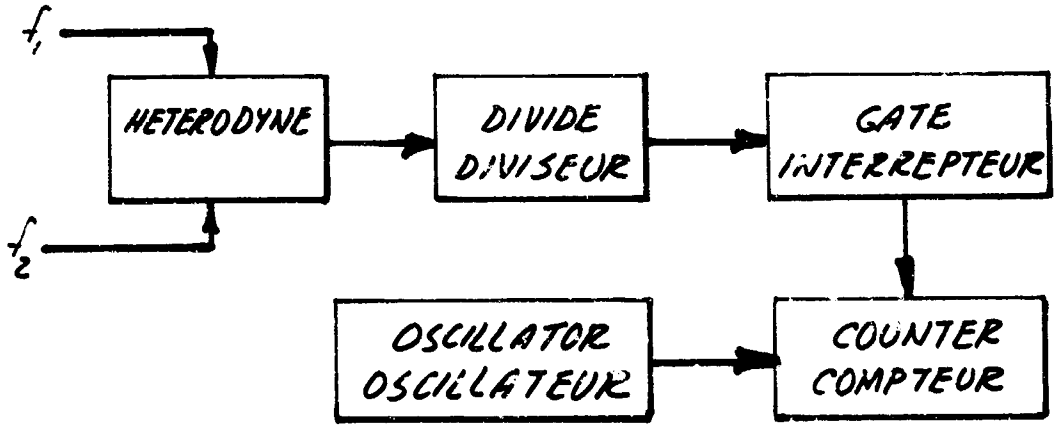
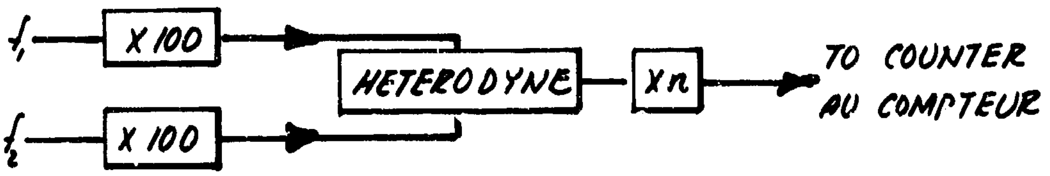


Fig. 5 Readout method No. 2
Méthode de présentation No. 2



FREQUENCY MULTIPLIER - MULTIPLICATEUR DES FRÉQUENCES

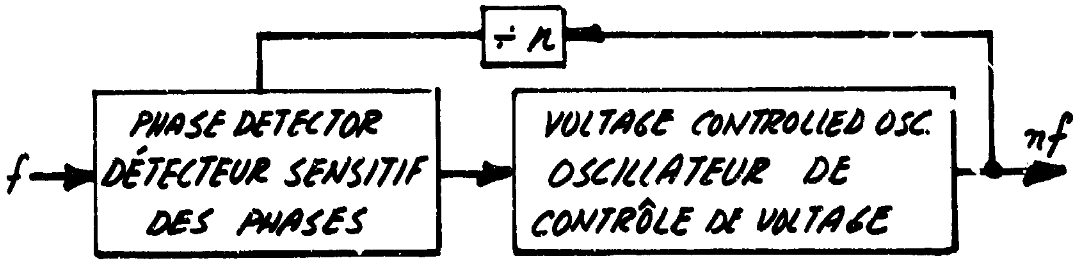


Fig. 6 Readout method No. 3
Méthode de présentation No. 3

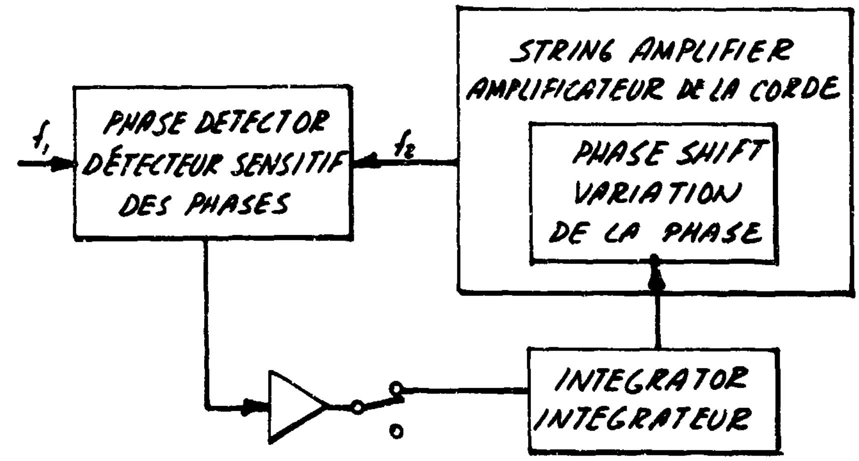


Fig. 7 Automatic in-orbit zero adjust
Réglage automatique du zéro

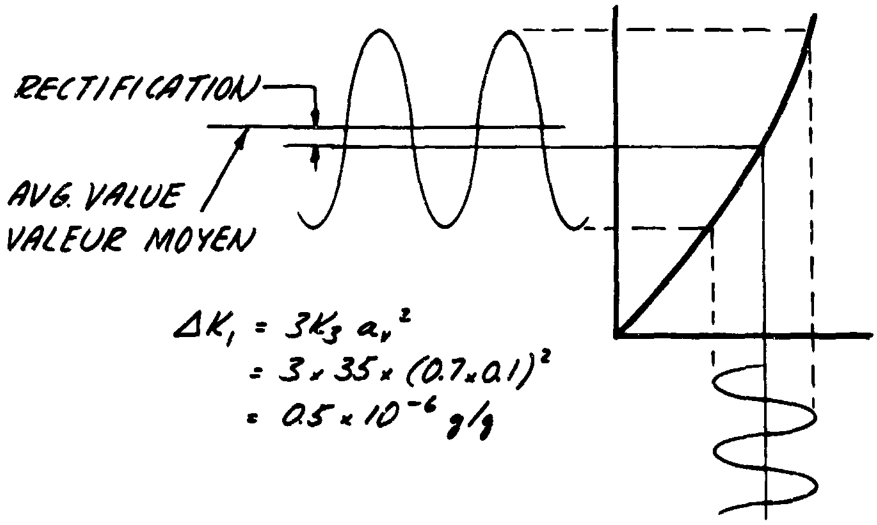

RECTIFICATION
AVG. VALUE
VALEUR MOYEN
$\Delta K_1 = 3K_3\, a_v^2$
$= 3 \times 35 \times (0.7 \times 0.1)^2$
$= 0.5 \times 10^{-6}\, g/g$

Fig. 8 Rectification effect
Effet de rectification

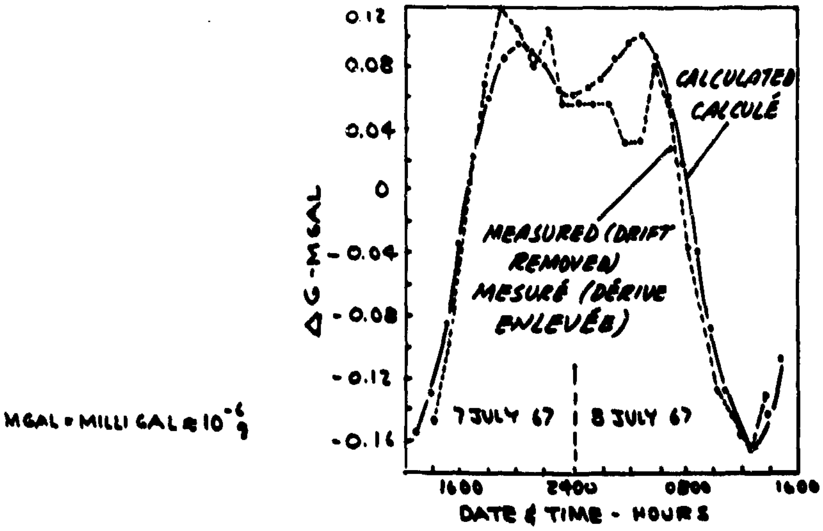


Fig. 9 Lunar tide measurement
Mesurage de la marée de la lune

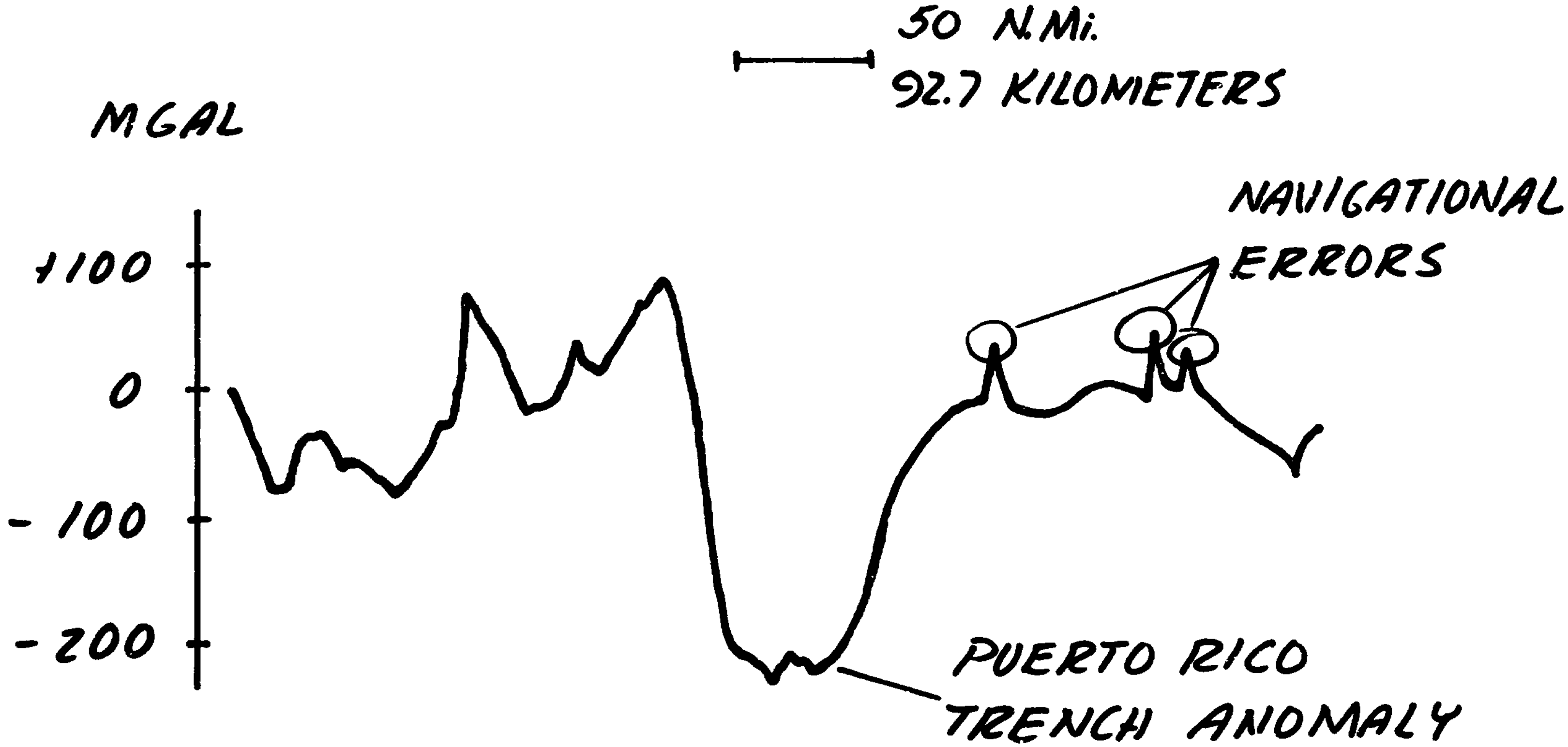


Fig. 10 Free-air anomaly (drift not removed). RV chain cruise 75

TECHNICAL REPORT ON THE
QUARTZ RESONATOR DIGITAL ACCELEROMETER

by

Norman R. Serra

Litton Systems, Inc.
Woodland Hills, California, USA

SUMMARY

The Quartz Resonator Digital Accelerometer is the result of an extensive program to develop an accelerometer with digital output using direct conversion from force to frequency. Crystalline quartz, as the basic transducer material, provides high mechanical stability, repeatability, and reliability and enables direct mechanical-to-electrical coupling using the piezoelectric properties of the material. This eliminates conventional electrical pickoffs and torquers usually associated with precision accelerometers. Instrument accuracy is built into the quartz resonator. Relative simplicity of the required electrical circuitry eliminates the need for critical circuit components. The crystals of two quartz resonator-mass systems operating in push-pull control the frequency of two oscillator circuits. The obtained difference frequency is proportional to the applied acceleration and provides an extremely accurate digital output with no requirement for intermediate equipment.

The initial phase of the program was devoted to development of the quartz resonator. A six-degree-of-freedom computer program was used to establish dimensions and to calculate resonant frequencies and harmonics. The importance of resonator activity level was established and direct measurement techniques developed. Effects of cutting, lapping, polishing, etching, scribing techniques and aging on performance and stability were experimentally determined. The instrument design effort required development of a temperature-insensitive proof mass and cross support system, a torsion free cantilever suspension, a stable, distortion-free case, a rigid external mounting system, and an ultra-high vacuum seal.

The instrument, in its final configuration, consisted of a dual crystal, dual proof mass design with each proof mass suspended by a single cantilever spring. The electronics consisted of two independent crystal controlled circuits, with each crystal incorporated into the bridge of a high stability oscillator. Multiplier and frequency differencing circuits were developed for direct instrument readout.

The design goal for the instrument was 0.02 percent linearity in the 0 to 20 g range. This was achieved and demonstrated in centrifuge tests at the US Army Guidance and Control Laboratory at Huntsville, Alabama.

The program made significant contributions to the state-of-the-art in the development of analytical expressions for frequency vs load, in elimination of mechanical lock-in of the crystals, in developing techniques for fabrication, tuning, matching, and mounting of crystals, in the study of single mode crystal operation, and in the development of precision crystal-controlled, variable frequency circuits with megohm range, series, resonant crystal impedances.

TECHNICAL REPORT ON THE
QUARTZ RESONATOR DIGITAL ACCELEROMETER

Norman R. Serra

1. INTRODUCTION

1.1 Digital Accelerometers

The primary requirements for linear digital accelerometers for inertial navigation systems and guided missile applications are simplicity, low cost, small size, reliability, and accuracy. Previous approaches to high performance designs have involved the use of pulsed forcing, or analog to digital conversion of force-balance analog instruments. The pulsed systems require tightly controlled voltage supplies and precision wave shaping to obtain accurate digital output. Analog to digital conversion offers the disadvantages of extra equipment requirements and loss of accuracy at the conversion stage. Recently, the use of g-sensitive vibrating elements with open-loop outputs has been tried with a significant degree of success. The Litton quartz crystal resonator is of this type.

Open-loop digital instruments have the advantages of simplicity, reliability, the elimination of pulsed forcing or A to D conversion, and have the potential for low cost, high rate production. While these instruments present challenging problems in the area of stability and temperature sensitivity, they offer excellent probability of meeting performance requirements in the areas of threshold sensitivity, dynamic range, and linearity.

1.2 The Litton Quartz Crystal Resonator

The open-loop digital transducer used in this instrument is a double, closed-ended, quartz tuning fork, as illustrated in Figure 1. It is approximately three-quarters of an inch in length, cut from crystalline quartz, and plated with thin metallic electrodes by means of which its beams are piezoelectrically excited at their mechanical resonant frequency. The precise configuration and dimensions of the resonator resulted from theoretical considerations and mathematical analyses which are described in some detail later in this paper.

Two quartz resonators are used in each instrument. Each resonator is fastened to the instrument case and to an individually suspended proof mass in a manner such that acceleration components along its axis of sensitivity place one in tension and the other in compression. The resultant stresses cause precisely defined changes in the resonant frequencies of the two resonators.

1.3 Advantages of Quartz

Quartz is one of several extremely stable crystals which exhibit piezoelectric properties and is used, by virtue of these properties, in frequency stabilization circuits for a multitude of applications. Bridge-stabilized crystal controlled oscillators commonly have frequency stabilities in the order of one part in 10^9. It is this extreme stability, coupled with a highly uniform and predictable change in frequency as a function of axial loading, that led to their choice for this application. A great amount of work in crystal controlled oscillators has been carried out by the electronics industry and, therefore, there is a large storehouse of background experience and information readily available. A further advantage of quartz is its anisotropic structure which allows its characteristics, such as Q , thermal coefficient of expansion, and frequency to be varied as functions of its cut.

1.4 Development Program Objectives

The primary objectives of the development program were to design, build, and test an instrument with the following design and performance goals.

(a) Digital Output:	Frequency as a function of acceleration.
(b) Simplicity:	Minimum number of parts.
(c) Low Cost:	Under $500 production cost.
(d) High Range:	Up to ±60 g if required.
(e) Reliability:	MTBF of 3000 hours or greater.
(f) Linearity:	0.02% or better.
(g) Stability:	0.1 milli-g.
(h) Low Cross Coupling:	100 μg per g in all axes.
(i) Temperature Operation:	-30^oC to $+80^oC$
(j) Small Size.	

The instrument was to be static tested at the Litton Industries facility at Woodland Hills, California, centrifuge tested at the US Naval Test Station at Point Mugu, California, and given additional tests at the Army Inertial Guidance and Control Laboratory at Huntsville, Alabama.

The main application of the instrument, if successfully developed, would be in inertial guidance systems for aircraft and in guidance and telemetering systems for missiles.

Additional objectives of the program were to evaluate the applicability of quartz crystals to this type of application, to develop procedures and techniques for fabrication of quartz tuning fork resonators, and to determine the producibility of these resonators for high quantity production situations.

2. RESONATOR DESIGN

2.1 Theoretical Considerations

A vibrating beam in axial tension or compression provides a simple direct technique for force to frequency conversion. A simple, conventionally mounted rigid beam, however, has serious deficiencies in sensitivity to end moments and mounting point energy losses. Incorporation of a pair of such beams into the symmetrical tuning fork configuration shown in Figure 2 serves to reduce this sensitivity greatly, and virtually eliminates transmission of kinetic energy to the mounting surfaces. The symmetrical construction allows the parallel beams to vibrate in symmetry, so that the moments due to bending at the dual beam junctions are cancelled. The compliant integral end supports provide a high degree of mechanical isolation between the active portion of the resonator and the fastening points at each end.

The differential equation for a fixed end beam of constant cross-section performing transverse vibrations was derived as shown in the Appendix. This equation is

$$EI\,\frac{\partial^4 Y}{\partial x^4} - T\,\frac{\partial^2 Y}{\partial x^2} + m\,\frac{\partial^2 Y}{\partial t^2} = 0 \;,$$

where

E is the modulus of elasticity in the x (axial) direction

I is the moment of inertia of a plane section through the beam

Y is the beam deflection at any point x

x is the position along the longitudinal axis of the beam

T is the axial load in tension

m is the mass per unit length

t is time.

Solution of this differential equation for fixed end boundary conditions results in a transcendental equation of the form:

$$\left(a_1\,\frac{l}{2}\right)\tanh\left(a_1\,\frac{l}{2}\right) + \left(a_2\,\frac{l}{2}\right)\tan\left(a_2\,\frac{l}{2}\right) = 0 \;,$$

where

l is the length of the beam

$$a_1 = \left[\sqrt{\left\{\left(\frac{T}{2EI}\right)^2 + \frac{m\omega^2}{EI}\right\}} - \frac{T}{EI}\right]^{\frac{1}{2}}$$

$$a_2 = \left[\sqrt{\left\{\left(\frac{T}{2EI}\right)^2 + \frac{m\omega^2}{EI}\right\}} - \frac{T}{2EI}\right]^{\frac{1}{2}}$$

ω = resonant frequency.

A computer program was used to solve for ω, resulting in the graph shown in Figure 3. This has the following series expansion about $T = 0$:

$$\omega = 5.59332\left(\frac{2}{l}\right)^2\sqrt{\left(\frac{EI}{m}\right)}\left[1 + 0.09831\left(\frac{l^2T}{8EI}\right) - 0.00531\left(\frac{l^2T}{8EI}\right)^2 + 0.00055\left(\frac{l^2T}{8EI}\right)^3 - \ldots\right] .$$

Details of the entire analysis are shown in the Appendix. This result shows that the equation for natural frequency as a function of acceleration for a resonator in tension is of the form

$$f_T = f_0 + k_1 a - k_2 a^2 + k_3 a^3 - \ldots \;,$$

where f_0 is the no-load natural frequency; and the equation for a resonator in compression is correspondingly

$$f_C = f_0 - k_1 a - k_2 a^2 - k_3 a^3 - \ldots \;.$$

These are the basic equations used in design of this instrument.

The properties of the various crystallographic orientations were determined on the basis of published data and calculations[1, 2] which define thermal coefficient of frequency and Q as a function of crystal cut. It was established that thermal coefficients of frequency can be reduced to zero for individual preselected temperatures, but that crystals cut for this characteristic do not provide maximum piezoelectric coupling. The cut finally used was chosen for the practical reasons of availability and high coupling.

2.2 Physical Description

The resonator, shown in Figures 1, 2, and 4, is symmetrically shaped in the form of a double closed-ended tuning fork. Its dimensions and tolerances in inches are 0.800 (±0.001) length, 0.080 (±0.0005) width, and 0.040 (±0.00005) thickness. The twin resonating beams in the center are 0.250 (±0.0001) inches in length by 0.010 (±0.0001) inches in width and the center section is connected to the end mounting areas by 0.015 (±0.001) inch wide end tines.

The material used was an optical grade, high-Q ultraviolet grade, cultured quartz. Natural quartz was also obtained for evaluation, based on the recommendation of Bell Telephone Laboratories, where natural quartz had been shown to provide higher Q. No conclusions as to relative merits for this application were reached.

The crystallographic orientation used in the final design is a Y cut, in which the face of the resonator is in a plane perpendicular to the Z axis of the crystal, with the length aligned along the X axis, and the width along the Y axis.

The electrode pattern is illustrated in Figure 5. This simple symmetrical pattern, with voltages applied as shown, causes the beams to vibrate in their primary mode. Piezoelectric deformation results from impressing electrical fields across the narrow dimension of the beam, which is along the crystallographic Y axis. With this mode of excitation, quartz distorts in shear, forcing the beams to bend in a transverse direction as shown, or in the opposite direction when the field is reversed. The beams vibrate at the frequency of the applied field, and these vibrations maintain maximum amplitude when that frequency coincides with the natural frequency of the beams.

The plated pattern is formed with a 500 to 1000 Ångstrom thickness of nickel-gold plating, scribed with a special tool to isolate the electrodes and their circuits mechanically and electrically. The circuitry to the electrodes is plated on the faces of the crystal adjacent to the electrodes. The detailed pattern of electrodes is also shown in Figure 4. Plated circuitry eliminates the need for conventional wiring connections on the resonator. The resonant frequencies of the resonators are 28 and 30 kHz, and two frequencies are used in order to avoid frequency lock-in, a problem which is discussed in more detail below. The 30 kHz range was chosen for the best compromise between mechanical and electrical characteristics.

3. INSTRUMENT DESIGN

3.1 Theoretical Approach

The equations for the natural frequencies of the resonator in tension and in compression have been derived. In this instrument, the resonators and their proof masses are oriented such that an acceleration along the sensitive axis places one resonator in tension and the other in compression. Taking the difference of the two resonator frequencies, and dropping higher order terms,

$$\begin{aligned} f_t - f_c &= f_0 + k_1a - k_2a^2 + k_3a^3 - f_0 + k_1a + k_2a^2 + k_3a^3 \\ &= 2k_1a + 2k_3a^3 \end{aligned}$$

gives an expression which is quite linear, since the second-order terms cancel, and the third-order terms are small (less than 0.02% at 20 g in the prototype units). Thus it is seen that the frequency differencing system provides outputs which are proportional to applied acceleration.

The use of a two resonator system, shown schematically in Figure 6, provides advantages beyond the reduction of non-linearities. The f_0 , or no-load resonant frequency terms of the two resonators, are subtracted from one another in the frequency difference equation. It can be seen that, if the changes which occur in no-load resonant frequency as a function of environment and time are equal in direction and magnitude for both resonators, their effects will also mutually cancel.

Each proof mass is supported by a simple cantilever spring, which lies in a plane perpendicular to the longitudinal axis of the crystal and to the sensitive axis of the instrument. This provides cross-support in the two directions perpendicular to the sensitive axis. The spring constant of the support is calculated to provide temperature compensation. This is achieved by determining the amount of thermal differential expansion which occurs between the resonator and the case, and allowing the force applied to the resonator through the spring to be of a magnitude which will compensate precisely for the temperature coefficiency of frequency of the resonator under steady-state thermal conditions.

3.2 Physical Description

The instrument, shown in Figure 7, is $2\frac{7}{8}$ inches in diameter and 2 inches in length. It is constructed of two nearly identical instrument halves, bolted together at their mounting flanges, and each containing an independent resonator and proof mass assembly. A cut-away view is shown in Figure 8. Symmetry is achieved by locating each resonator in a hole in the center of its proof mass, with the two ends of the resonator attached to round mounting buttons, using epoxy cement. The buttons are, in turn, cemented to the case structure at one end of the resonator and to the proof mass at the other. The button at the case end contains two sealed feedthrough terminals which provide direct electrical connections from the outside of the case to the resonator.

The proof mass is a composite assembly, the center portion of which is an integral part of the spring. Each spring actually consists of two spring leaves in the same plane, laterally separated for optimum moment stiffness. The proof mass centers of gravity are located so that they become irrotational upon application of linear accelerations.

The instrument halves are sealed with a Viton O-ring and an indium metal gasket seal, in order to maintain the low leak rate (less than 10^{-12} std. cc/sec) required for long life. The use of organic cements and gaskets within the instrument limits the outgassing temperatures to under 100°C, but this is sufficient to yield predictable crystal performance. A $\frac{3}{8}$-inch diameter copper tube, soldered into one case half, serves as an evacuation, outgassing, and final pinch-off orifice.

3.3 Major Design Problems

The initial design of the instrument was intended to meet the functional requirements of scale factor, range, and linearity in a package of minimum size and weight. While tests of individual instrument halves indicated that the linearity goal could be met, complete instruments were plagued with the problems of frequency lock-in, case sensitivity to external forces, diminished resonator Q , and inadequate vacuum sealing.

Lock-in is caused by the two crystals resonating together in sympathetic vibration, apparently due to a mutual interchange of mechanical and acoustical energy. It was eliminated by operating the crystals at two different frequencies, with a frequency ratio of $\frac{13}{14}$. This gives a separation of approximately 2000 Hz, eliminating the possibility of lock-in over the entire acceleration range.

The first case design was considerably smaller than the final, and was of lightweight aluminum construction. A cross-sectional sketch is shown in Figure 9. Early tests revealed a sensitivity to mounting conditions that, in many instances, caused the resonators to vary no-load frequency, and occasionally to cease vibrating completely. Joining of the separate instrument halves caused similar problems, which could only be eliminated by repeated interchange of instrument halves until a satisfactory combination was found. This proved extremely difficult because of the epoxy sealing of the instrument halves which had to be completed to obtain an integral joint.

A new case was designed in the configuration described in Section 3.2 and shown in Figures 7 and 8, which greatly stiffened the resonator supporting structure and eliminated mounting point distortions. An O-ring seal was included to permit ease of assembly and disassembly during laboratory investigation of the mechanical-acoustic energy transmission problem. The planes of the two proof mass cross-support systems were brought into line so that they would be at the same radius during centrifuge testing. The original design of the resonator mounting, proof mass, and cross-supports was not changed in the new "heavy case" design.

When resonators were installed in each half of the heavy case accelerometer, a vast improvement in Q over the earlier design was exhibited. However, when the two halves were united, a 50% drop in resonator activity was experienced. The assembled halves presented a large area of metal to metal contact which permitted a substantial interchange of energy from resonator to resonator. This area was reduced by placing thin metal shims between the halves at six points around the flange. The measured drop in resonator Q was now reduced to less than 25%.

In order to insure a final seal, an indium wire gasket was added between the case halves. This material has the property of deforming under pressure to accept the contour of the surfaces with which it is in contact, filling up small surface imperfections and maintaining its volumetric integrity over an indefinite period.

4. ELECTRICAL DESIGN

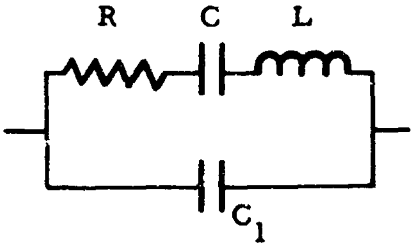
4.1 Theoretical Approach

The fundamental problem of resonant frequency vibration is the design of a circuit which will tune itself precisely to the natural frequency of the controlling element (in this case, the resonator). For this purpose, each crystal was incorporated into a bridge oscillator circuit designed to operate with the unusually high series impedance of the quartz crystal resonator. The circuit schematic is shown in Figure 10.

The general considerations that apply to any oscillator are expressed by Barkhausen's criteria which require that, while in a steady state condition, the power gain around a closed loop be unity and that the phase shift be zero. If gain exceeds unity, oscillations will build up in amplitude until limiting occurs in the loop, thus reducing the gain. The second of these criteria, which requires that the phase shift around the loop be zero, determines the operating frequency. For a non-zero phase shift condition, the frequency changes the proper amount to cause the phase shift to become zero. The circuit design features used to satisfy those criteria are described in Section 4.2.1.

In addition, the resonator stability must be defined in useful and measurable terms. The first order mode of the crystal is shown below, where C_1 is the equivalent shunt capacitance of the crystal, and L , C , and R represent the motional impedance.

The figure of merit M is a reliable measure of crystal stability,



where

$$M = Q\frac{C}{C_1}$$

$$Q = \frac{1}{\omega_0 RC} .$$

In conventional applications, M is normally greater than 10, but in this application it ranges from 0.3 to 3.0. Its importance is shown in the frequency stability expression

$$\frac{\Delta f}{f} = -\frac{2\left(1 + \frac{1}{M^2}\right)}{Q(\mu + 2)} d\phi ,$$

where

$\Delta f/f$	is the fractional change in frequency
μ	is the amplifier net voltage gain
$d\phi$	is the phase change in the circuitry external to the bridge.

For a typical crystal

$$Q = 2 \times 10^4$$

$$M = 1$$

$$\mu = 200$$

$$d\phi = 0.05$$

$$\Delta f/f \simeq 5 \times 10^{-8} .$$

4.2 Circuit Design

4.2.1 Excitation Circuit

A large variety of oscillator circuits have been designed for use with low to moderately high impedance crystals. One such circuit is the Meacham bridge oscillator[3], which was designed at Bell Laboratories and served to demonstrate that it is possible to minimize the frequency changes due to electronic shifts external to the bridge itself. In that circuit, the reactance of the shunt capacitance relative to the series impedance of the crystal is relatively high. The crystal appears as a small resistance shunted by a large capacitance. A transformer is used to obtain the bridge drive, and a thermal element is used to control amplitude.

The Litton quartz resonator, however, electrically resembles a large resistor shunted by a capacitor with comparable reactance. The series resonant impedance is in the $\frac{1}{2}$ to 6 megohm range. For this reason, a more complex circuit was required. This circuit, shown in Figure 11, uses two driving sources 180 degrees out-of-phase with one another instead of a transformer as used by Meacham. This provides a broader available frequency band and permits control of one of the two driving amplifier loops with an AGC. The Litton circuit uses this AGC rather than a thermal element for amplitude control. With amplitude tightly controlled, the frequency becomes effectively dependent upon the phase shift. The extremely high impedance pickoff from the bridge minimizes the bridge loading and the resultant effects upon amplitude and phase shift. The loops including the oscillators have extremely high gain, in excess of 100. This also has a stabilizing effect, by tending to reduce the effects of phase shift due to electronics outside the bridge, as shown by μ in the equations above.

With these stabilizing factors, the only remaining area of serious concern is the stability of the RC balancing leg of the bridge. As shown above, the effect of the external electronics on the frequency stability is in the neighborhood of five parts in 10^8.

4.2.2 Output Electronics

The output mechanization, shown in Figure 12, consists of a $\frac{15}{16}$ multiplication of the higher frequency crystal output, an additional 5× multiplication of both outputs, and a frequency differencing circuit. The basic technique for multiplying the frequency of a signal is to extract an appropriate harmonic after altering the signal waveform to assure a known spectrum. Odd harmonic multiplication (×15) is performed by square-wave limiting of the input signal to be multiplied initially and using appropriate filtering to extract the fifth and then the third harmonic, with square-wave limiting as an intermediate step. The resulting output is a frequency 15 times that of the input. Division by 16 is achieved with a four stage binary counter. Multiplication by 5 is the same as multiplication by 15 with the elimination of the last two steps.

A standard mixer circuit is used to obtain the difference frequency between the two resonator outputs. None of its components is critical and, with the use of high frequency silicon transistors, special precautions in layout and packaging are unnecessary.

5. RESONATOR FABRICATION

5.1 Procedures

Several essential sequential operations were involved in the resonator manufacturing process: ultrasonic cutting of the resonator from a single blank, lapping and polishing to match the beams and improve activity, intermediate testing, plating and scribing, final testing and matching in pairs, and operational aging at temperature for optimum stability. Development of the sequence, tooling, and technique was a lengthy and highly empirical effort which ultimately resulted in a predictable and acceptable yield of crystals.

5.1.1 Resonator Cutting

A Sheffield Cavitron was used for ultrasonic cutting of the resonators from flat quartz blanks of desired finished thickness. The goal in this area was to develop tooling and methods for cutting symmetrical resonators, with closely repeatable dimensions, at an acceptable production rate. In addition, the crystal surfaces were to be substantially free from cracks, chips, and other defects.

A three-piece cutting tool was designed which allowed the critical cutting surfaces of the tool to be replaced readily without major tool rework, and which cut as many as seven resonators consecutively from a flat plate without the necessity of sizing each blank. Brass tools were used to eliminate the danger of chipping. However, satisfactory results and better tool wear were ultimately achieved with steel. A slurry of 3 micron Al_2O_3 was used for cooling, lubrication, and removal of waste material.

This equipment produces a resonator in less than one hour of machine time with negligible breakage. The raw-cut resonators are of good quality and surface finish, and tolerances on the critical dimensions are controlled to within 100 microinches. While further improvements are possible, these methods yield acceptable raw-cut resonators at reasonable cost.

5.1.2 Lapping and Polishing

The natural frequency of the resonator is highly dependent upon the dimensions of the active beams of the tuning fork. A change of 100 microinches in thickness can produce a 300 Hz change in natural frequency; therefore, frequency variations due to machining must be removed by more precise techniques. Frequency trimming by adding mass to the beams by overplating was unsuccessful because results were too limited to compensate for the large variation between raw-cut beams. Fine grit sandblasting by which frequencies could be controlled down to 5 to 10 Hz was also attempted, but it was not as effective as lapping and polishing, which achieves 1-Hz control. Etching in a solution of hydrofluoric acid was found to yield a great improvement in beam activity. The apparent mechanism is the removal of surface irregularities and small particles of debris remaining after vigorous cleaning and lapping. The final procedure established was to lap to within 100 Hz of the desired frequency and then polish and etch to the final frequency.

5.1.3 Intermediate Testing

Frequency matching of the beams of an individual resonator is an indispensable requirement for high-Q performance. Matching of the resonators destined for a particular instrument is also highly desirable from the standpoint of linearity, temperature compensation, and aging. Since the only acceptable way to match frequencies is by removal of material from the surface of the active beams, tuning and matching can only be accomplished on unplated resonators. Because of this, it is necessary to measure frequency at the unplated polishing stage.

A capacitive drive fixture consisting of eight flat plates arranged in a pattern similar to the electrode pattern of the resonator was designed to perform this measurement. The beams of the unplated crystal were located between the fixture plates with a small physical clearance to assure that no mechanical contact was made, and voltages were applied in a manner identical to the resonator electrode pattern. Natural frequency and Q are readily determinable from the voltage output measurements.

5.1.4 Beam Balancing of Resonators

An important consideration in the performance of this transducer is its sensitivity to stimuli other than forces along its sensitive axis. It is desirable to mount the resonator in a manner that prevents the application of non-axial forces; but, since this is difficult to accomplish, the resonator must be made as insensitive as possible to the cross-axis forces imposed by distortions in the clamps and the proof mass support system. Analysis indicates that frequency changes due to side loads in a plane which includes the two active beams can be minimized by insuring that the two active beams are dimensionally balanced. This was examined experimentally by clamping one end of a plated resonator in a stationary frame and applying a series of cross-axis loads at the other end. The graph of frequency versus side loads plotted in Figure 13 was obtained from these data. A second resonator which had been "beam balanced" was examined under the same circumstances. The data, plotted in Figure 14, show it to be less sensitive to side loads by a factor greater than 10. Beam balancing by polishing and etching was therefore established as a mandatory step in the resonator fabrication process.

5.1.5 Plating and Scribing

Application of the electrodes and circuitry to the resonator in a manner which would provide reliable adhesion, uniform thickness, homogeneous structural qualities, and freedom from surface discontinuities involved many serious problems, the foremost of which was the choice of materials. Best results were normally obtained with bimetallic plating, the underlay material being chosen for surface adhesion and the overlay for uniformity, structure, and surface qualities. Several materials and combinations, aluminum, chromium and gold, nickel and gold, platinum, and others were tried, using resonator performance as a measure of acceptability. A nickel-gold combination gave the best empirical results and was used in the prototype instruments. The plating was performed with conventional vacuum deposition equipment using a rotational fixture to obtain uniform plating.

The two obvious methods of obtaining electrode and circuitry pattern definition are to mask the paths between circuits prior to plating, or to use photo-etching to remove material after plating, a process similar to techniques used in printed microcircuitry. Unfortunately the internal surfaces of the beams do not lend themselves well to these techniques and, consequently, a method of direct material removal, using an electrical scribing tool, was devised instead. With this method, manual skill and care in handling are the prime requisites for success, and these were developed to the extent that the operation became routine. It seems apparent, however, that further experimentation with the previously described approaches might yield considerable savings in time and improved consistency in results. The electrode and circuitry pattern are illustrated in Figures 4 and 5.

5.1.6 Resonator Aging

Because of the relationship between output Δf per g and the no-load natural frequency of the resonators, it can be seen that the no-load natural frequencies must be extremely stable (in the range of a fraction of a part per million per month). Thus, one of the major technical endeavors in the program was to improve stability by controls applied to the fabrication of the resonators, and by specially developed techniques designed to accelerate physical stabilization of the crystals. The latter is referred to as aging, and consists of high temperature continuous operation of the resonators until such time as initial large changes in no-load frequency have disappeared.

The aging process is normally carried out in a vacuum oven controlled to 0.01 degree Centigrade and is aided by constant high vacuum pumping for continuous outgassing of the crystal. A typical aging curve is shown in Figure 15. It can be seen that the greatest aging occurs in the first 400 hours, after which the average slope of the aging curve becomes more gradual. It also should be noted that aging of the electronic components in the oscillator circuits was found to be occurring simultaneously, and that their relative contribution to the composite aging rate becomes large as the crystal itself becomes more stable.

Best results were obtained when the aging process was applied to completed instruments, held at temperature, with evacuation occurring through the copper fill tube, permitting the unit to be sealed off without further contamination by the outside atmosphere. Stabilities obtained were in the order of one part in 10^7 random drift.

6. PERFORMANCE

6.1 Type of Testing Conducted

The tests performed on the instrument and its component parts were those which could yield information particularly germane to the problems of this specific design. Among the most critical problems were stability versus time and stability versus temperature of both the crystal-instrument combination and the electronic oscillator. These characteristics were tested extensively. Other variables tested because of their critical effect upon performance were the level of activity of the crystal as a function of acceleration, the sensitivity of the instrument to changes in internal pressure, and the oscillator stability as a function of the a.c. drive voltage of the crystal and of the d.c. power level to the oscillator. Finally the instrument was subjected to conventional tests for zero to one g linearity and centrifuge linearity to ±20 g.

6.1.1 Stability versus Time

A large number of stability tests were conducted on individual resonators, one of which is shown in Figure 15. The ultimate interest, however, is focused on resonator performance as a part of the completed instrument, since it is essential to know how long an aging period is needed for the instrument to stabilize, and what level of stabilization can be expected. For this purpose, the heavy case instrument was placed in a temperature controlled environment and operated for a five-month period. Output was monitored on a continuous basis and plotted daily. The results are shown graphically in Figure 16. It was observed that it took almost the full five months for the instrument to achieve an acceptable degree of stability and that, even then, a small residual drift uncertainty, in the range of 0.1 milli-g, persisted. The test was discontinued after experiments showed that the drift rate could be varied by minor component changes in the electronics, from which it was concluded that the area of electronic component stability would have to be investigated and improved before additional meaningful drift tests could be conducted.

Although the initial drift of the instrument was high (0.1 mill-g per day), the progressive and relatively predictable decrease in drift rate during the test period indicated a probability of achieving ultimate stability in the fractional milli-g range by control of resonator processing and supplementary aging procedures.

6.1.2 Stability versus Temperature

There has not as yet been an adequate checkout of the original thermal design theory of the instrument, which specified that the axial loading produced by differential expansion between the case assembly and resonators and transmitted to the resonators by the spring should cancel the thermally induced frequency shifts in the resonators. There were, however, tests performed on individual resonators for frequency versus temperature, which can be helpful in predicting thermal performance. Figure 17 shows a typical resonator curve. It can be seen that a preferred temperature exists at which minimal temperature sensitivity occurs and that there is a precise point at which the df/dt becomes zero. The rate of change at 60°C, a typical operating temperature, is 0.7 parts per million per degree Centigrade.

An examination of oscillator characteristics under similar conditions (shown in Figure 18) indicated a well defined temperature characteristic which will also have a measurable effect on the overall thermal stability. The net thermal sensitivity due to the sum of all thermal error contributions is as yet unknown; however, it was observed that control of temperature to 0.05°C during test appeared to be sufficient to eliminate significant errors.

6.2 Accelerometer Parameters

6.2.1 Resonator Activity Level

It was demonstrated early in the program that the first requisite for repeatable instrument performance was the maintenance of a high level of activity, or Q , on both resonators. Two factors which were originally suspected of contributing to a reduction in activity level were the level of internal pressure within the instrument, and the actual application of g-loads along the sensitive axis. Both of these were tested and the results are shown in Figures 19 and 20. The pressure versus Q curve shows that internal pressure can be increased to 10^{-1} torr without seriously impairing performance. Acceleration, fortunately, had no measurable effect on resonator Q .

6.2.2 Oscillator Circuit Performance Errors

Two critical variables, the power supply voltage to the oscillator and the crystal drive level, were also observed to have a measurable effect on the instrument output period or frequency. Test results, shown in Figures 21 and 22, demonstrated this to be a problem requiring controls and tolerances on both variables. The power supply must be held to ±0.1 volts and the crystal drive level to ±1% to limit the errors from these sources to one micro-g or less.

6.2.3 *Zero to One g Linearity*

One g linearity testing is conducted by rotating the sensitive axis of the instrument in 10 degree increments through 360 degrees in a vertical plane through the earth's gravity fields. Readings of accelerometer output frequency are made at each increment of rotation. The data are reduced to obtain the straight line scale factor coefficient and bias, and the residual error at each point is plotted as a percent of 1 g. Figure 23 shows the test results on the heavy case completed instrument. Maximum non-linearity on this particular test is 0.0013% of 1 g, and is typical of 0 to 1 g tests on both heavy and lightweight case designs.

6.2.4 *Zero to 20 g Linearity*

Multi-point centrifuge testing was conducted at two different government facilities. In the tests at Point Mugu, California, a set of ten test points was used, while at Huntsville, Alabama, a series of repeatability tests was conducted at 0, ±10, and ±20 g. The 10-point test data are plotted in Figure 24, which shows percent non-linearity with the resonator equations for frequency versus load. The maximum non-linearity was 0.02% in the positive direction and 0.035% in the negative, both of which occurred at the maximum g level.

The 0, 10, and 20-g repeatability tests showed a scattering of data ranging from 0.006 to 0.02% non-linearity. The scatter was due to poor centrifuge speed control, but the magnitudes served to confirm and even improve upon the 10-point test results, and verify that the expected non-linearity is in the 0.02% range at the 20 g full scale output.

7. CONCLUSIONS

7.1 Major Results and Conclusions

The major results and conclusions listed below summarize the accomplishments of the Quartz Resonator Digital Accelerometer development program.

(a) The practicability of using a quartz resonator, closed-end tuning fork as an open-loop acceleration sensing device has been demonstrated.

(b) The practicability of a resonator-oscillator electronic mechanization which will accurately represent, with respect to frequency, the internal stress levels within the resonator has been established.

(c) The effectiveness of a frequency-differencing system, which serves to eliminate inherent non-linearities in the resonator force-output relationships, has been demonstrated.

(d) The major problems inherent in the manufacturing of quartz crystal elements for the precise and unusual requirements of this type of application have been shown to be surmountable.

(e) Both the mechanical and electrical performance of this type of system have proven to be subject to mathematical analysis which, it was demonstrated, can be sufficiently rigorous to predict performance characteristics and problem areas accurately.

(f) The accelerometer developed during this program demonstrated performance which is reasonably satisfactory in range, linearity, threshold, and output format, but is not yet completely acceptable from a functional instrument standpoint. This performance does, however, serve well to demonstrate the principles involved, and tends to support the assumption that the remaining problems are subject to reasonable engineering solution.

7.2 Remaining Problems

The three areas upon which attention must be focused in order to attain universality of application for this instrument are stability, temperature sensitivity, and size. Although the instrument stability improved greatly over the five-month test period, an aging time of five months is unacceptable unless recalibration is permitted, and the level of stability achieved was only marginal. No serious attempt at stabilizing the electronic components has yet been made. Work in this area, as well as development of accelerated aging techniques for the resonators, is essential.

Temperature sensitivity has not yet been investigated thoroughly. There are several design approaches to decreasing the temperature sensitivity, if necessary, such as rotation of the crystallographic axes, variations in compliance of the cross-axis spring support, alteration of differential expansion rates by changing case assembly materials, thermal coefficient matching of the resonators, or attempting to operate at the minimal sensitivity temperature of both resonators.

Finally, the present size of the heavy case was chosen as a deliberate structural overdesign to solve the problem of mounting sensitivity. It remains to be seen how much the size can be reduced without re-initiating the problem.

ACKNOWLEDGEMENTS

Sincere appreciation is due to several individuals who made indispensable contributions to the accomplishments of this program, and to many other persons who assisted in a variety of ways. Their reports, analyses, sketches, and correspondence were used freely. I wish to express my sincere thanks to: H.P.Erdley, the inventor of the Litton Quartz Resonator and the conceptual designer of the instrument, and to R.A.McCann, R.A.Hansen, R.E.Stewart, S.A.Lever, R.W.Maughmer, M.Tarasevich, R.Levine, P.A.Norris, A.G.Grushkin, C.Meyers and J.G.Russell.

We are also highly indebted to those government agencies which provided interest and support, the US Army Inertial Guidance and Control Laboratory at Huntsville, Alabama, and the Air Force Missile Development Center at Holloman Air Force Base, New Mexico.

REFERENCES

1. Mason, W.P.	*Piezoelectric Crystals and Their Application to Ultrasonics.* Van Nostrand, 1950.
2. Beckman, R. et al.	*Higher Order Temperature Coefficients of the Elastic Stiffness and Compliances of Alpha Quartz.* Proceedings, Institute of Radio Engineers, Vol.50, August 1962, pp.1812-1822.
3. Meacham, L.A.	*The Bridge Stabilized Oscillator.* Bell Systems Technical Journal, Vol.XVII, 1938, pp.574-591.
4. Timoshenko, S. MacCullough, G.H.	*Problems in Strength of Materials.* Van Nostrand, 1948, p.163.

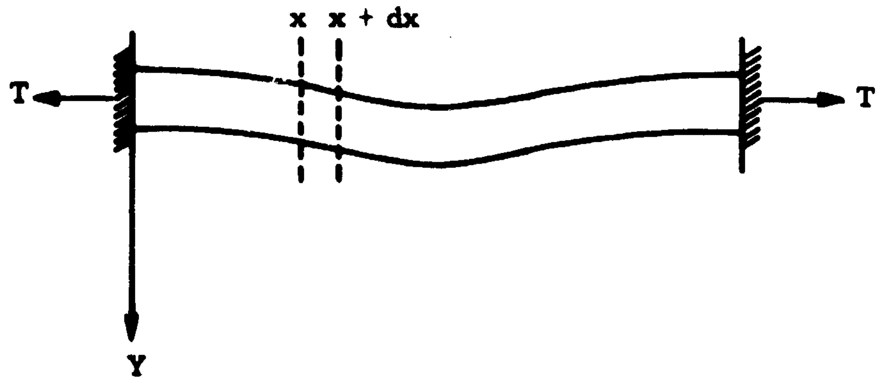
APPENDIX

DESIGN THEORY OF QUARTZ RESONATOR FORCE-TO-FREQUENCY TRANSDUCER

A.1 FREQUENCY RESPONSE OF A RESONATOR VIBRATING IN THE FUNDAMENTAL MODE

In this Appendix, the differential equation of a beam performing transverse vibrations is derived and then solved with the appropriate boundary conditions. Finally, the natural frequency of vibration is given as a function of the beam constants and the applied axial load.

A.1.1 Derivation of the Differential Equation



Consider a beam under an axial load (T) as shown in the sketch above; it is constrained at the ends but is free to vibrate in the transverse or Y direction at intermediate points.

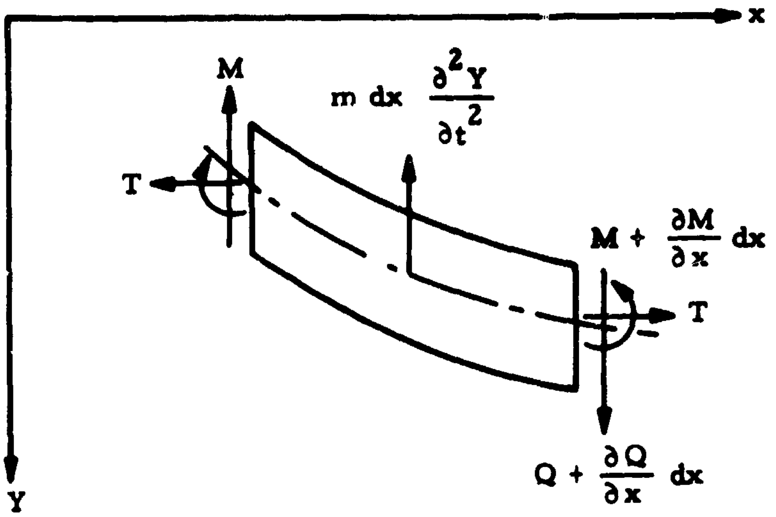
The differential equation may be derived by considering a small beam element of length dx at some position x . The diagram opposite shows all of the forces acting on the differential element where

T is the axial load

Q is the shear

M is the moment

m is the mass per unit length.


x
Y
M
$m\,dx\,\frac{\partial^2 Y}{\partial t^2}$
T
$M + \frac{\partial M}{\partial x}dx$
T
$Q + \frac{\partial Q}{\partial x}dx$

For equilibrium, the summation of all moments and forces, including inertial reaction, must equal zero. Summing moments about the center of gravity of the differential element gives

$$M + \frac{\partial M}{\partial x}dx - M - \left(Q + \frac{\partial Q}{\partial x}dx\right)\frac{1}{2}dx - Q\frac{1}{2}dx + 2T\frac{\partial Y}{\partial x}\frac{1}{2}dx = 0 ,$$

or, neglecting terms of second order,

$$\frac{\partial M}{\partial x} + T\frac{\partial Y}{\partial x} - Z = 0 . \tag{A1}$$

Summing forces in the Y direction gives

$$Q + \frac{\partial Q}{\partial x}dx - Q = m dx\,\frac{\partial^2 Y}{\partial t^2}$$

or

$$\frac{\partial Q}{\partial x} = m\frac{\partial^2 Y}{\partial t^2} . \tag{A2}$$

It can be shown[4] that, for small deflections the curvature and moment are related according to

$$M = -EI\frac{\partial^2 Y}{\partial x^2} , \tag{A3}$$

where

E is the modulus of elasticity in the x direction

I is the moment of inertia of a plane section through the beam.

Combining Equations (A1), (A2), and (A3) leads directly to the differential equation of a beam performing small transverse vibrations

$$\frac{\partial^2}{\partial x^2}\left(EI\frac{\partial^2 Y}{\partial x^2}\right) - \frac{\partial}{\partial x}\left(T\frac{\partial Y}{\partial x}\right) + m\frac{\partial^2 Y}{\partial t^2} = 0 . \tag{A4}$$

For the case of a beam of constant cross-section, Equation (A4) reduces to

$$EI \frac{\partial^4 Y}{\partial x^4} - T \frac{\partial^2 Y}{\partial x^2} + m \frac{\partial^2 Y}{\partial t^2} = 0 . \qquad (A5)$$

A.1.2 Solution of the Differential Equation

When the beam vibrates in one of its normal modes, the deflection at any position and time may be represented as

$$Y(x,t) = X(x) \sin \omega t . \qquad (A6)$$

Substituting Equation (A6) into Equation (A5) leads to

$$EI \frac{\partial^4 X}{\partial x^4} - T \frac{\partial^2 X}{\partial x^2} - m\omega^2 X = 0 . \qquad (A7)$$

It can easily be verified that $\cosh \alpha_1 x$ and $\cos \alpha_1 x$ are particular solutions of Equation (A7), where

$$\alpha_1 = \left[\sqrt{\left\{\left(\frac{T}{2EI}\right)^2 + \frac{m\omega^2}{EI}\right\}} + \frac{T}{2EI}\right]^{\frac{1}{2}} \qquad (A8a)$$

$$\alpha_2 = \left[\sqrt{\left\{\left(\frac{T}{2EI}\right)^2 + \frac{m\omega^2}{EI}\right\}} - \frac{T}{2EI}\right]^{\frac{1}{2}} . \qquad (A8b)$$

The general solution can thus be written as

$$X = A \cosh \alpha_1 x + B \cos \alpha_2 x , \qquad (A9)$$

where the constants A and B are chosen in accord with the boundary conditions. Let the origin be at the center of the beam. Since both ends are built-in, the boundary conditions are

$$X\left(-\frac{l}{2}\right) = X\left(\frac{l}{2}\right) = 0 \qquad (A10a)$$

$$\frac{dX\left(-\frac{l}{2}\right)}{dx} = \frac{dX\left(\frac{l}{2}\right)}{dx} = 0 . \qquad (A10b)$$

From Equation (A9), Equations (A10a) and (A10b) become

$$\left(\cosh \alpha_1 \frac{l}{2}\right) A + \left(\cos \alpha_2 \frac{l}{2}\right) B = 0$$

$$\left(\alpha_1 \sinh \alpha_1 \frac{l}{2}\right) A - \left(\alpha_2 \sin \alpha_2 \frac{l}{2}\right) B = 0 .$$

Since A and B are both non-zero, the determinant of the coefficients must vanish

$$\begin{vmatrix} \left(\cosh \alpha_1 \frac{l}{2}\right) & \left(\cos \alpha_2 \frac{l}{2}\right) \\ \left(\alpha_1 \sinh \alpha_1 \frac{l}{2}\right) & -\left(\alpha_2 \sin \alpha_2 \frac{l}{2}\right) \end{vmatrix} = 0$$

or

$$\left(\alpha_1 \frac{l}{2}\right) \tanh \left(\alpha_1 \frac{l}{2}\right) + \left(\alpha_2 \frac{l}{2}\right) \tan \left(\alpha_2 \frac{l}{2}\right) = 0 \, . \qquad \text{(A11)}$$

Equation (A11) is the transcendental equation which determines the relationship of frequency to the beam constants and the axial load.

A.1.3 Frequency Response

Because of the transcendental nature of Equation (A11) and the form of α_1 and α_2, an analytical solution for ω is somewhat tedious. However, the results of a computer program have yielded the graph shown in Figure 3 of the main text, which has the following series expansion about $T = 0$ or no load:

$$\omega = 5.59332\left(\frac{2}{l}\right)^2 \sqrt{\left(\frac{EI}{m}\right)} \left[1 + 0.09831 \left(\frac{l^2T}{8EI}\right) - 0.00531 \left(\frac{l^2T}{8EI}\right)^2 + 0.00055 \left(\frac{l^2T}{8EI}\right)^3 - \ldots\right] \, . \qquad \text{(A12)}$$

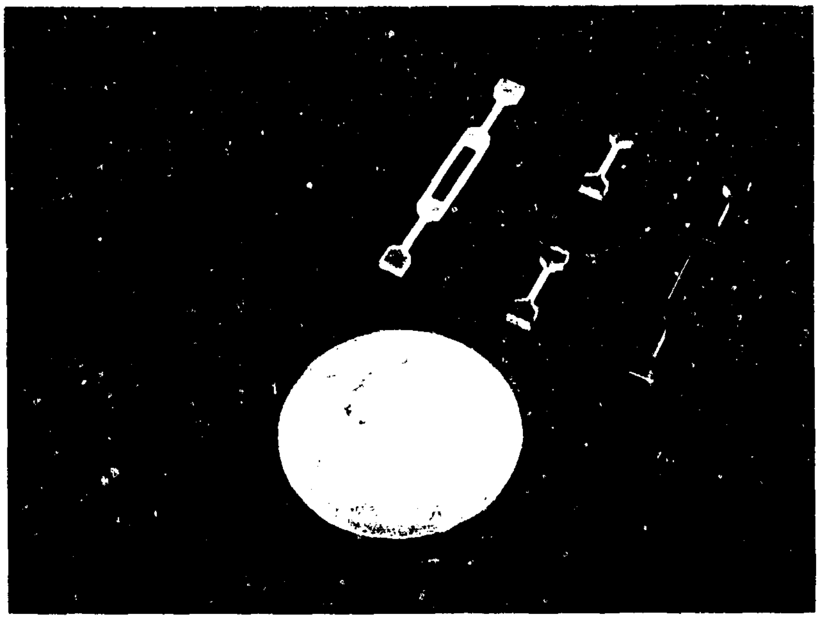


Fig.1 Quartz crystal resonator

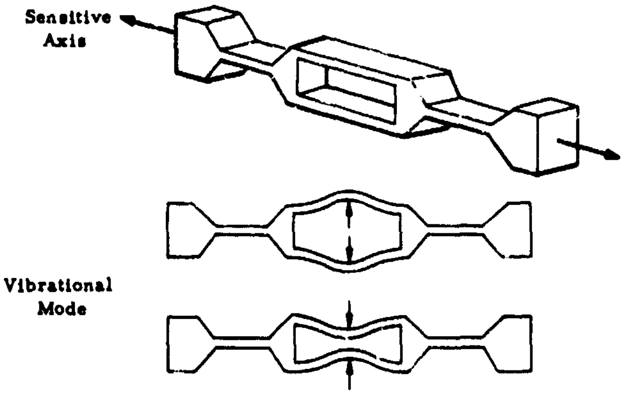


Fig.2 Sensitive axis and vibrational mode

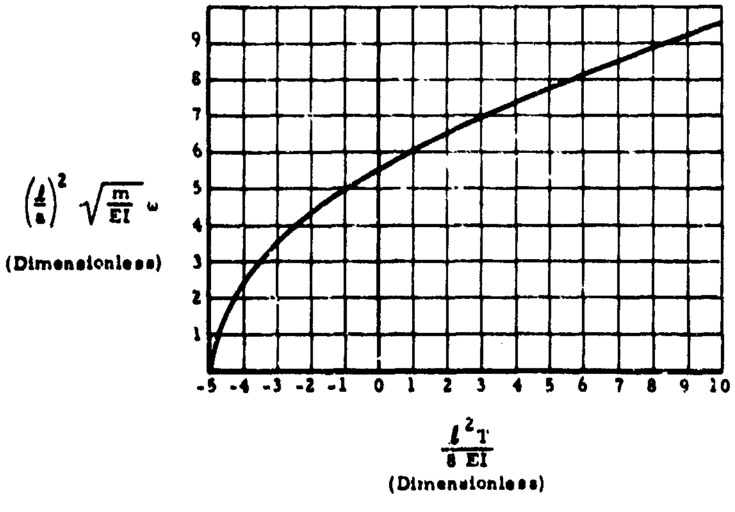


Fig.3 Frequency versus load response

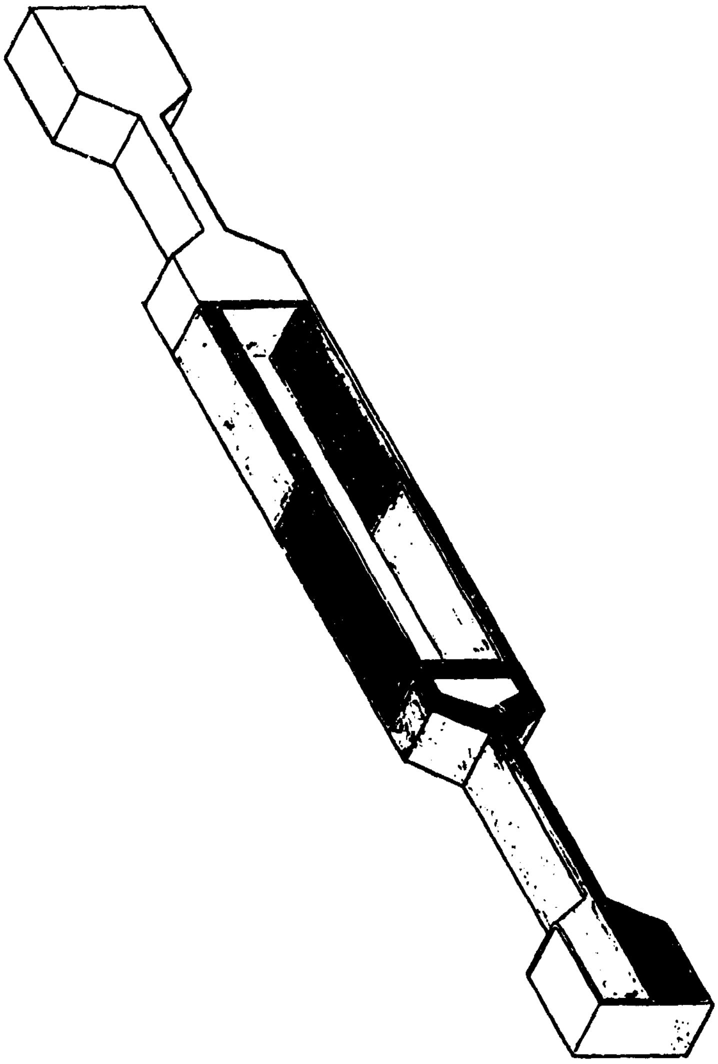


Fig.4 Plating pattern

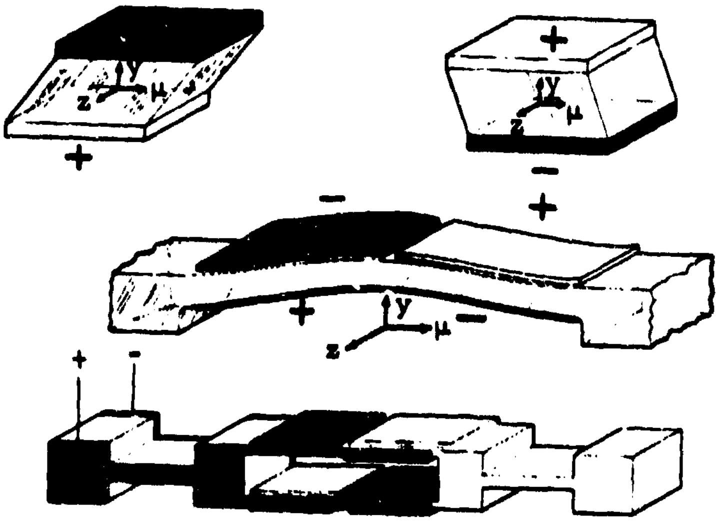


Fig.5 Piezoelectric excitation of oscillation

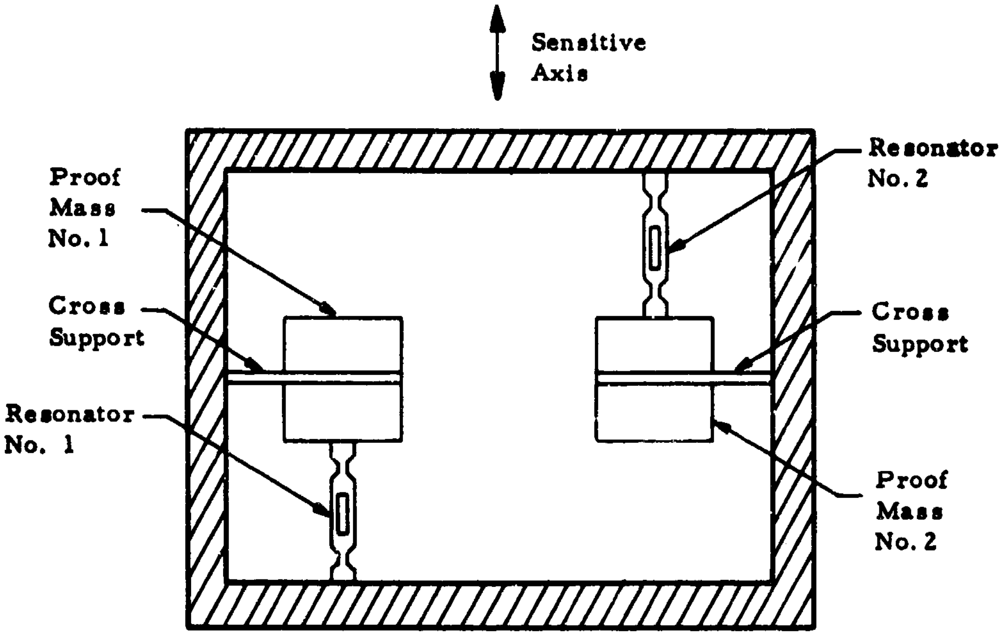


Fig.6 Schematic of two-resonator accelerometer

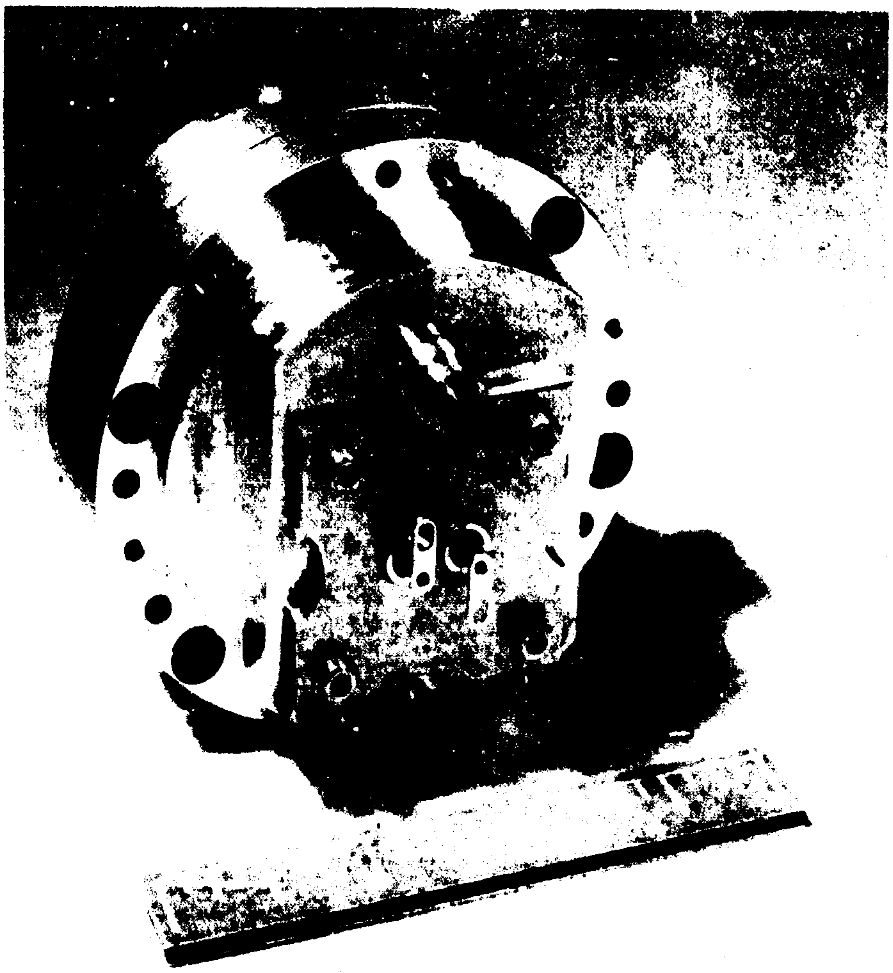


Fig.7 Heavy case accelerometer

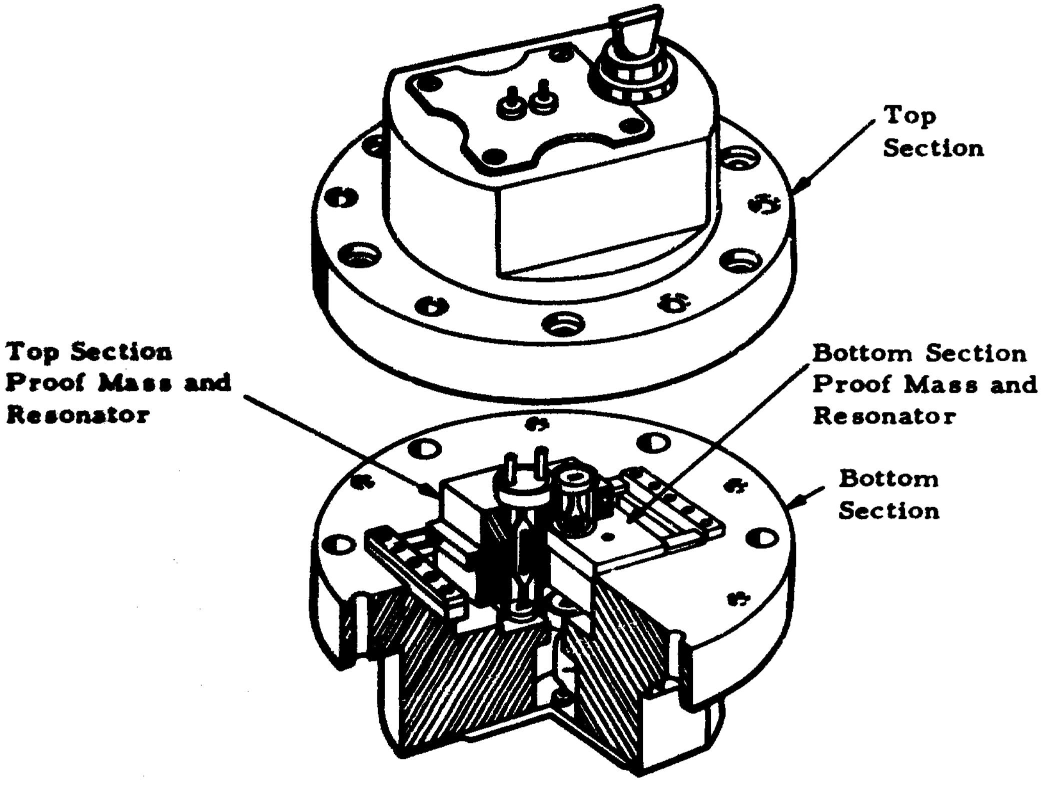

Top Section
Top Section Proof Mass and Resonator
Bottom Section Proof Mass and Resonator
Bottom Section

Fig.8 Final accelerometer design, cutaway view

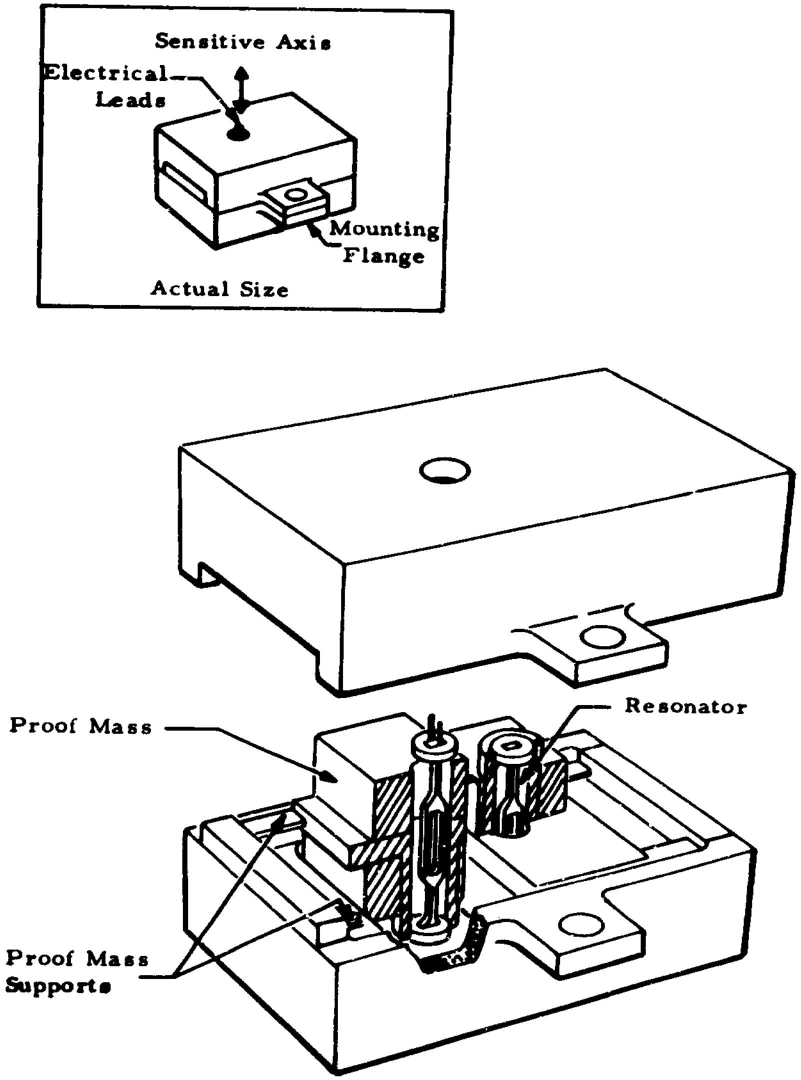
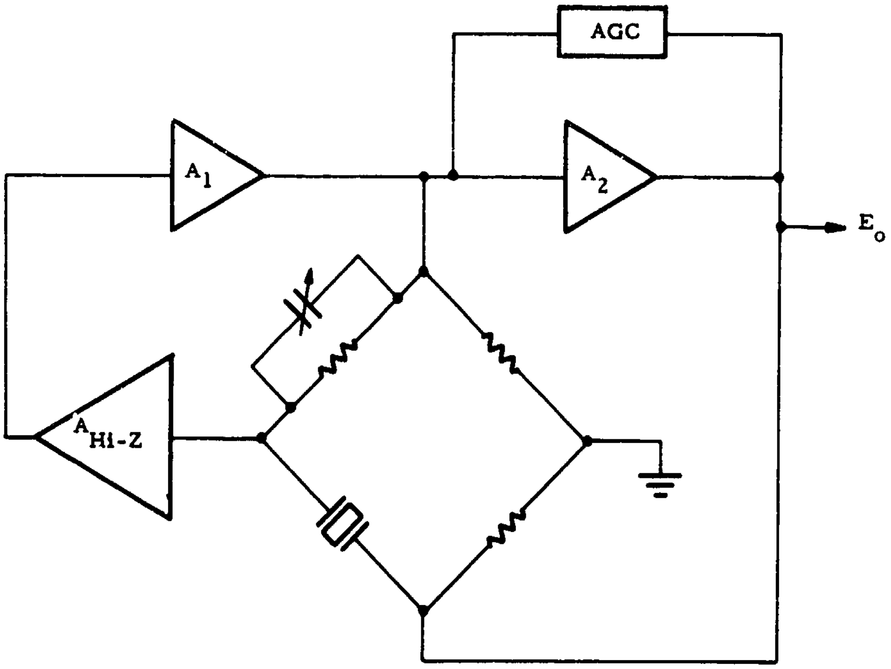


Fig.9 Original design, lightweight case



$A_2 = -A_1$

Fig.10 Bridge oscillator circuit schematic

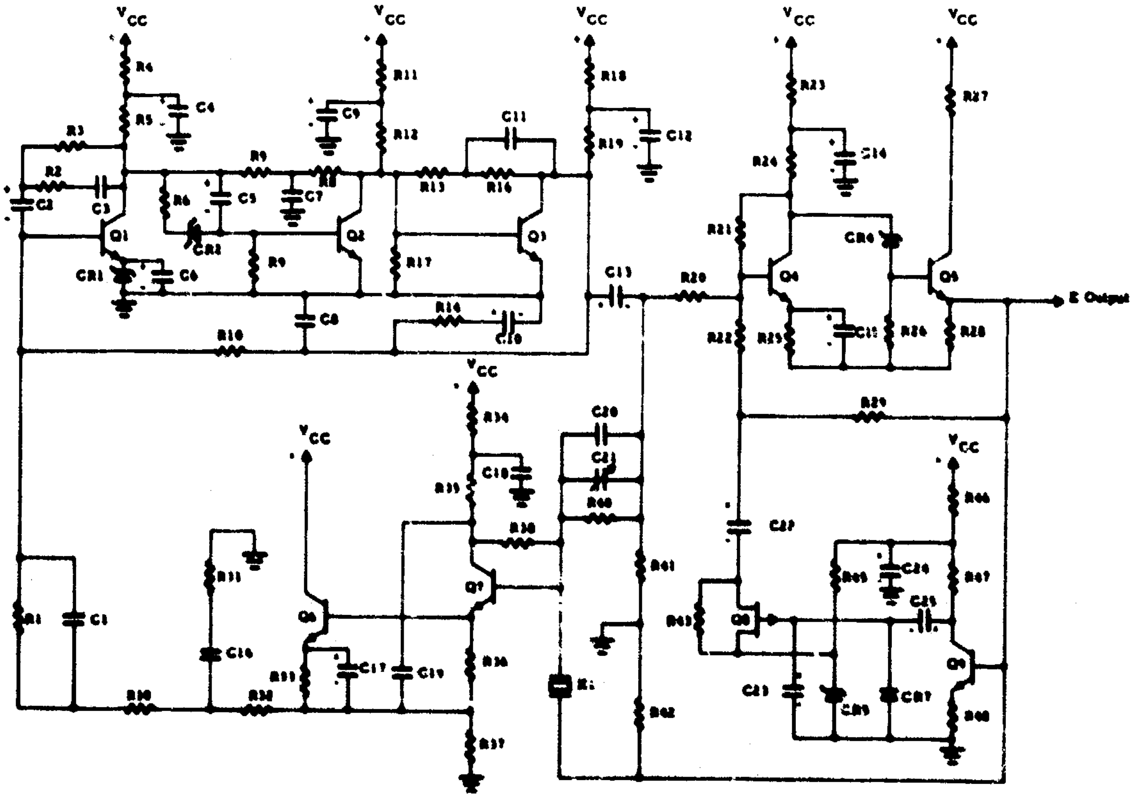


Fig.11 Oscillator circuit

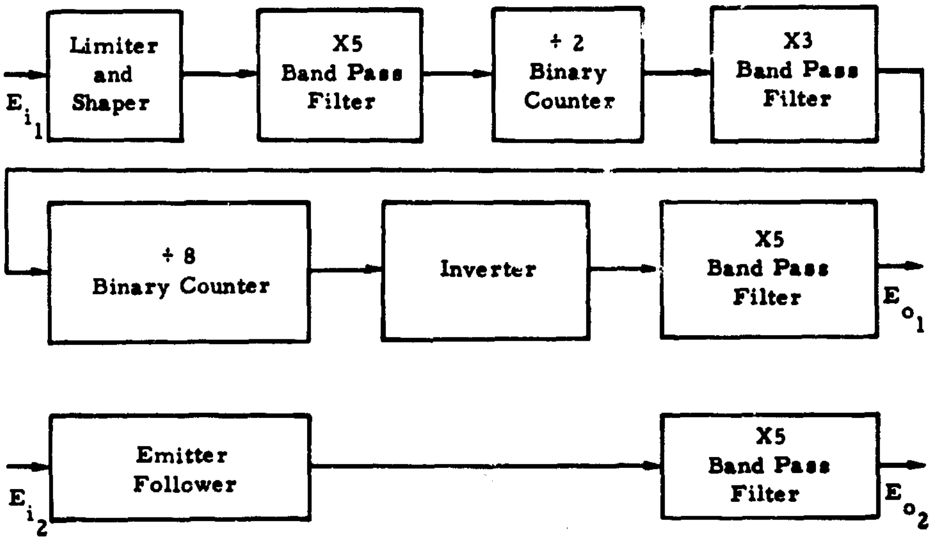


Fig. 12 Output mechanization

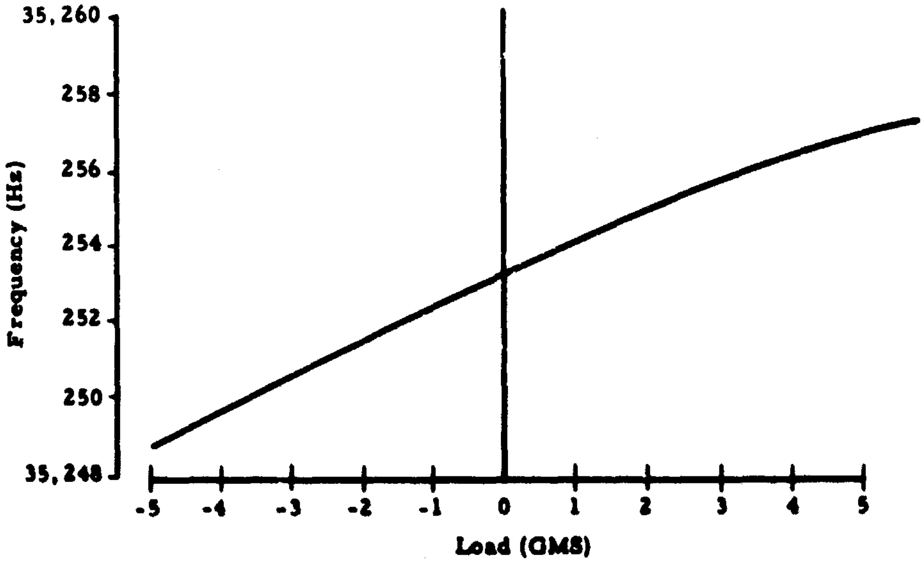


Fig. 13 Frequency versus side load (non-beam balanced)

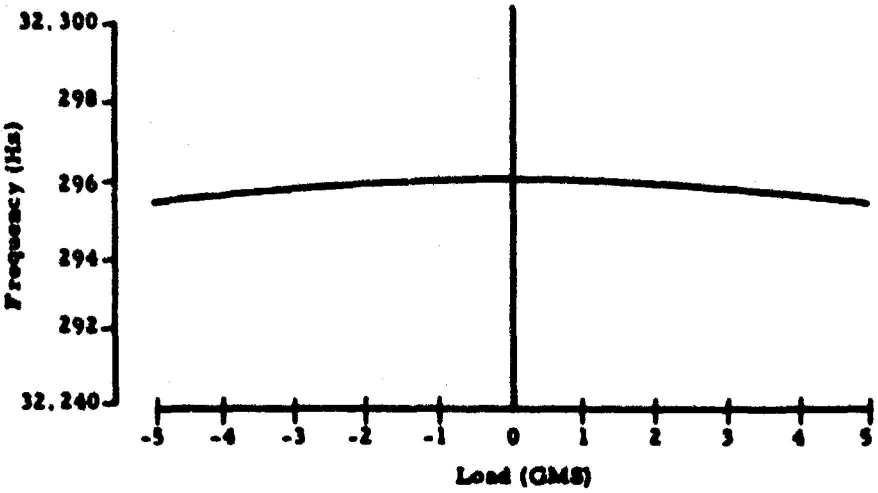


Fig. 14 Frequency versus side load (beam balanced)

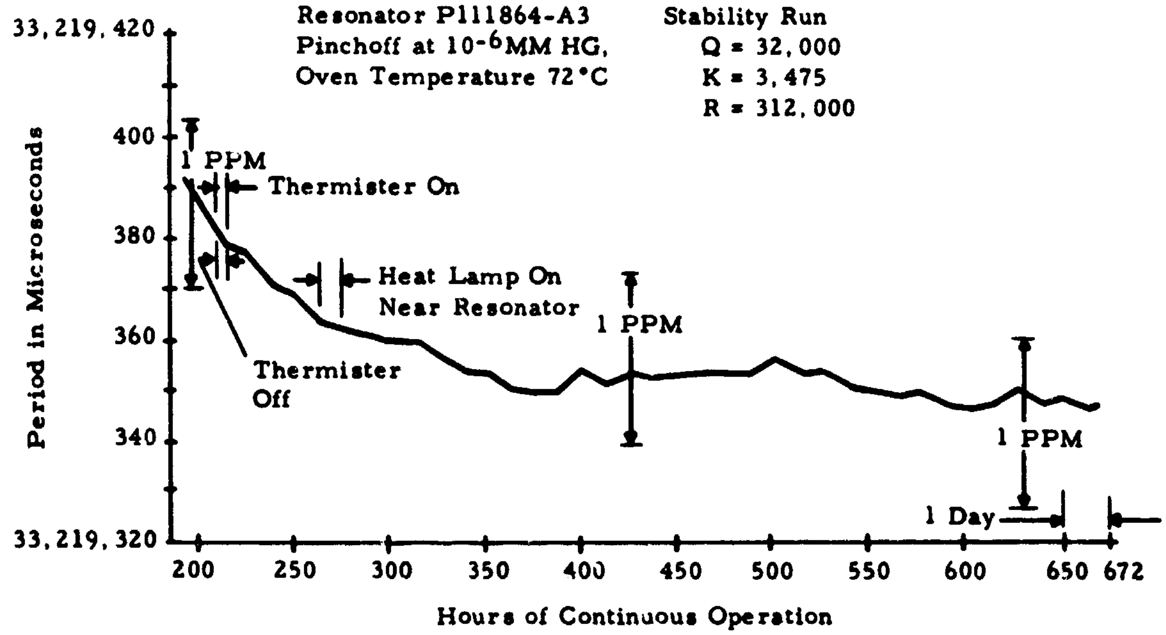


Fig. 15 Resonator stability

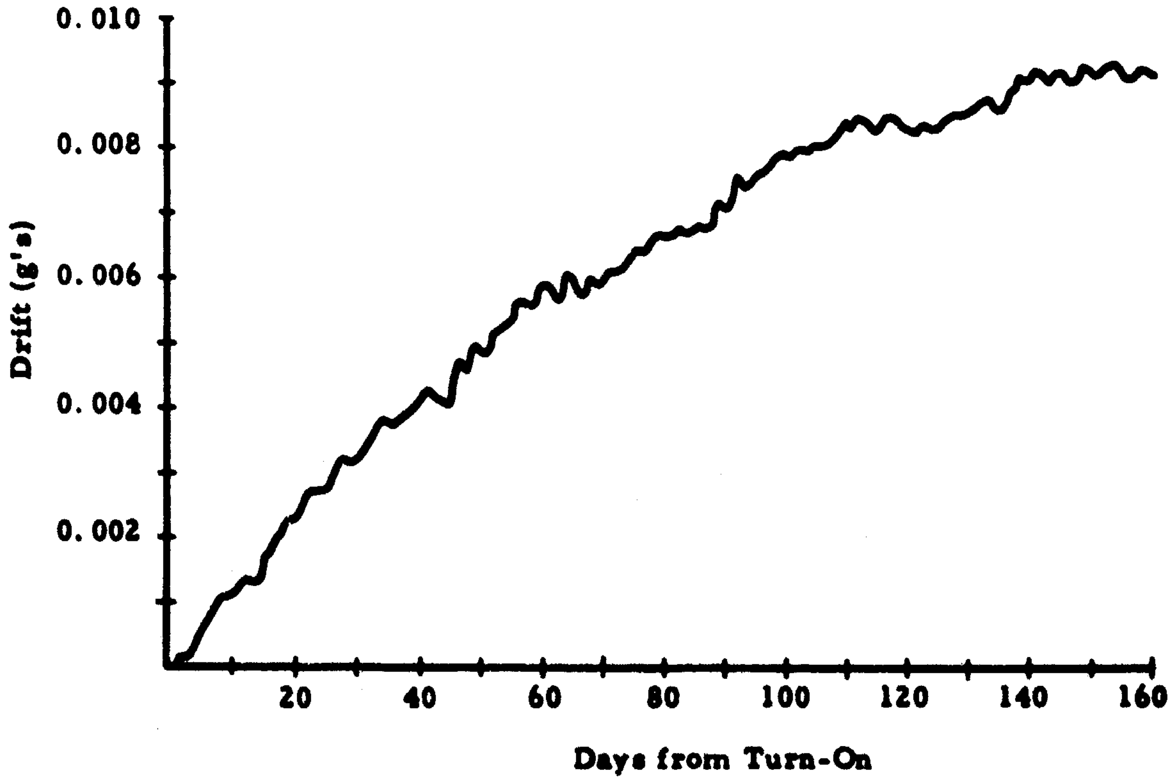


Fig. 16 Instrument stability

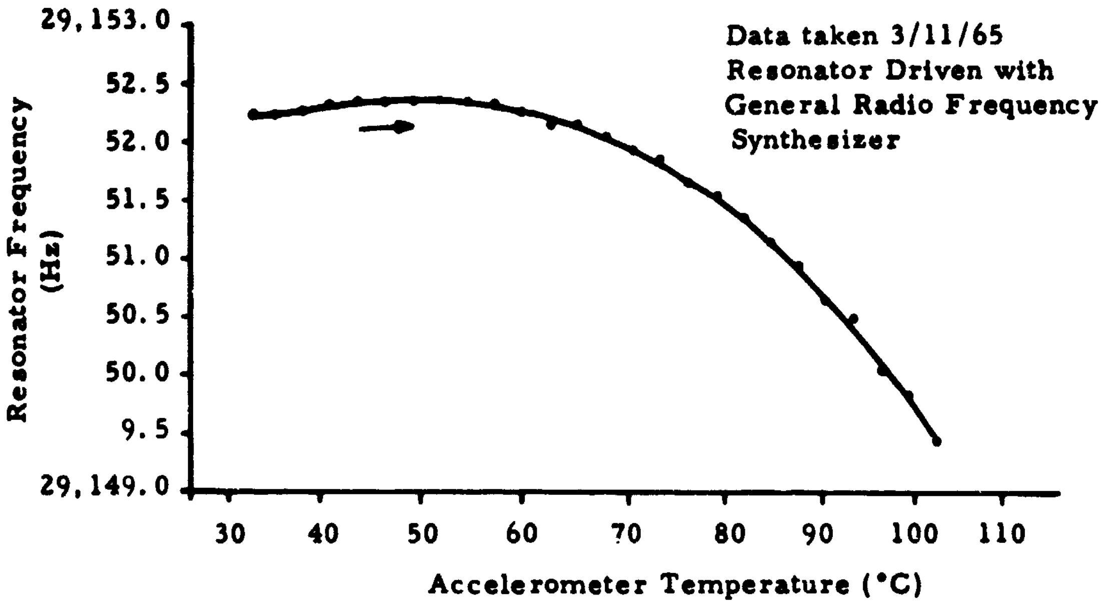


Fig.17 Frequency change versus temperature

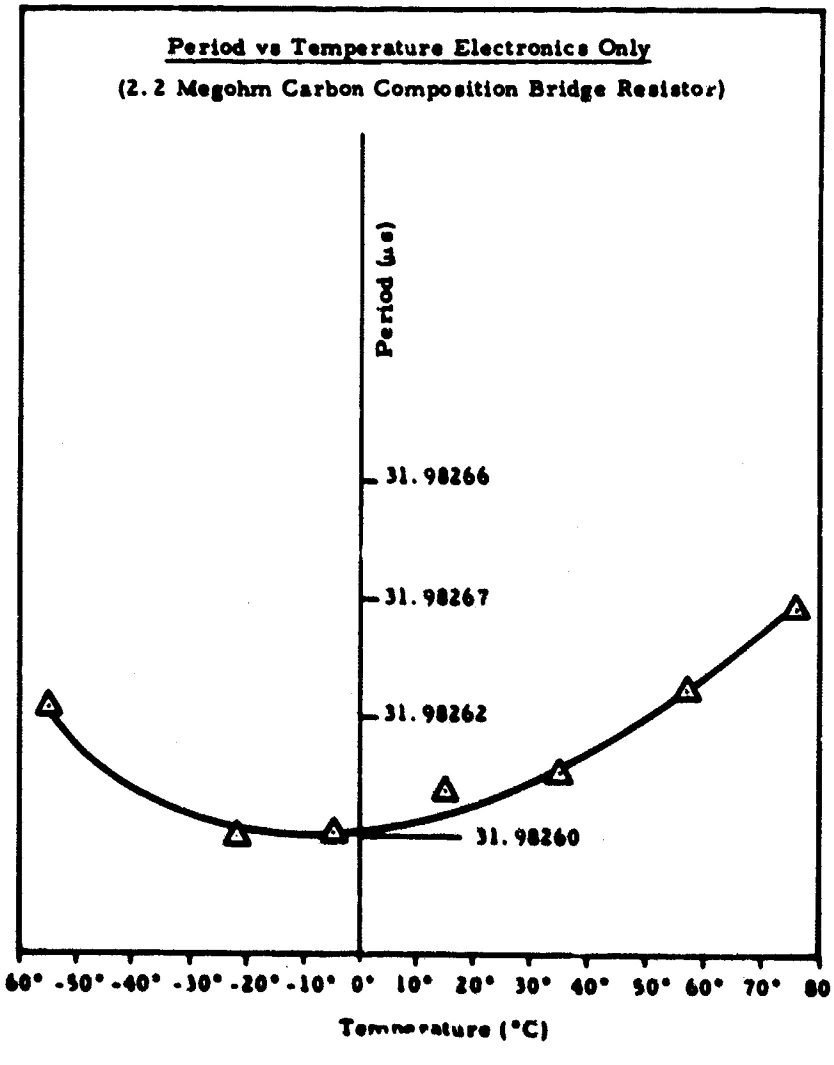


Fig.18 Oscillator output period versus temperature

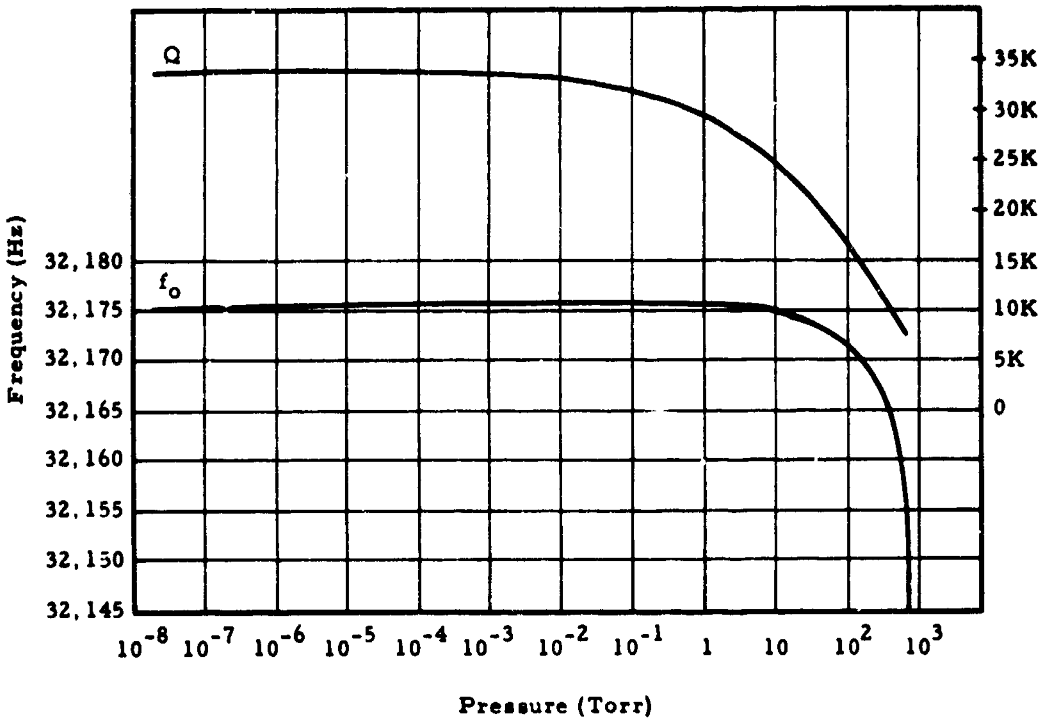


Fig.19 Resonator frequency versus pressure

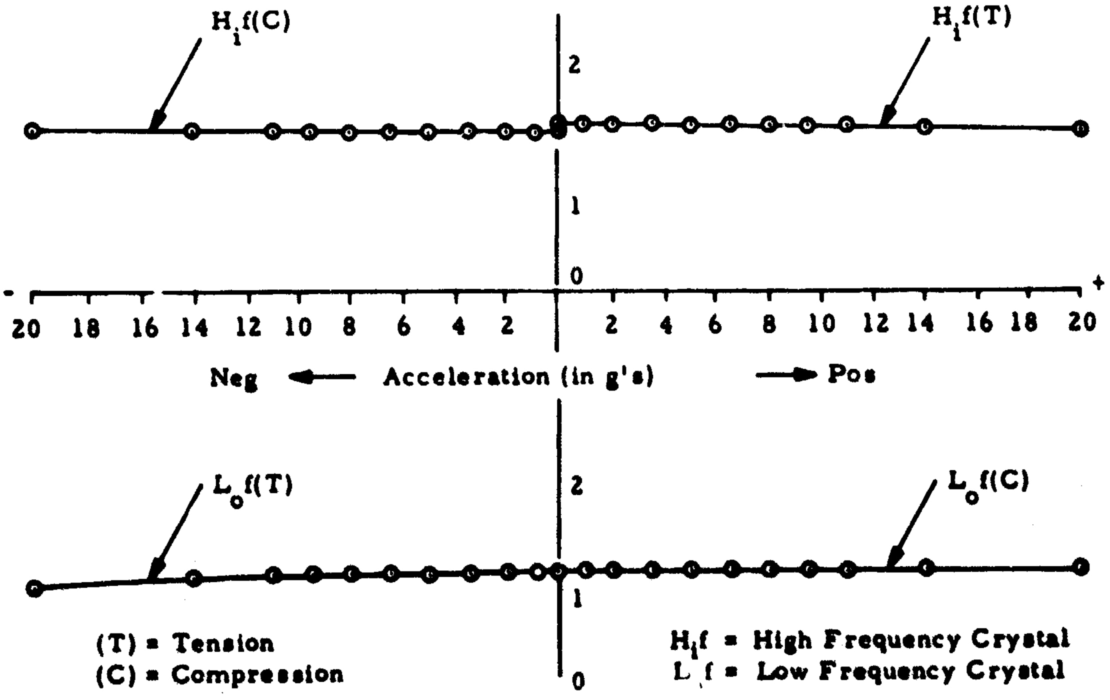


Fig.20 Resonator activity level versus acceleration

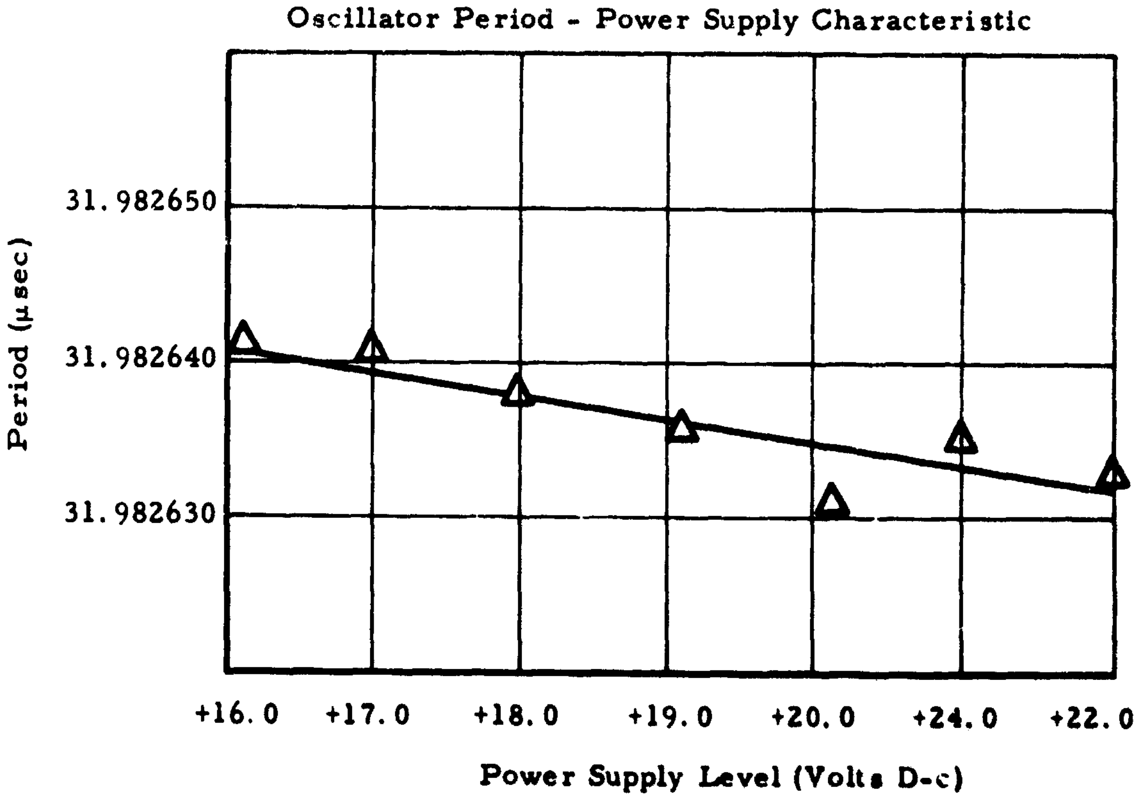


Fig.21 Oscillator output period versus supply voltage level

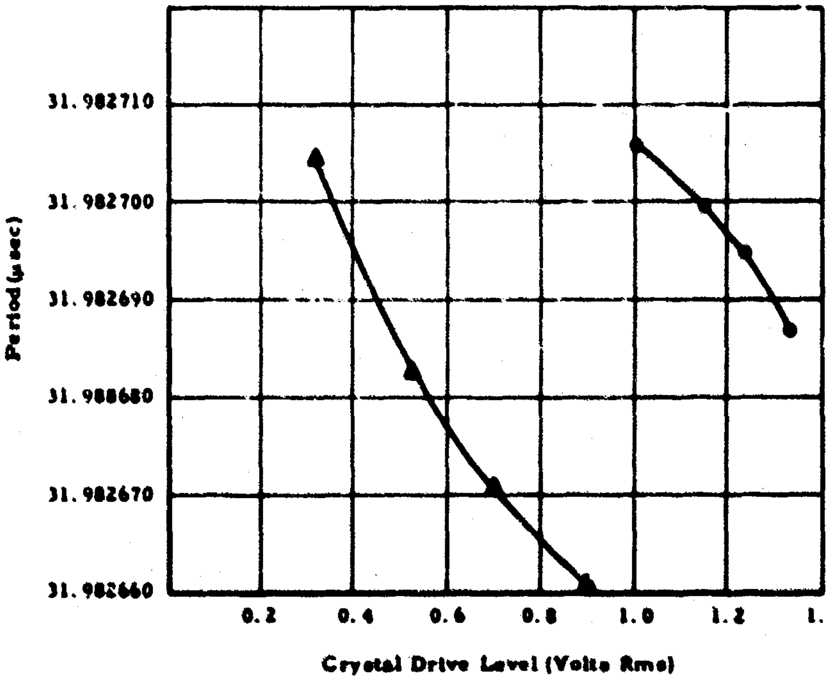


Fig.22 Oscillator period versus crystal drive level

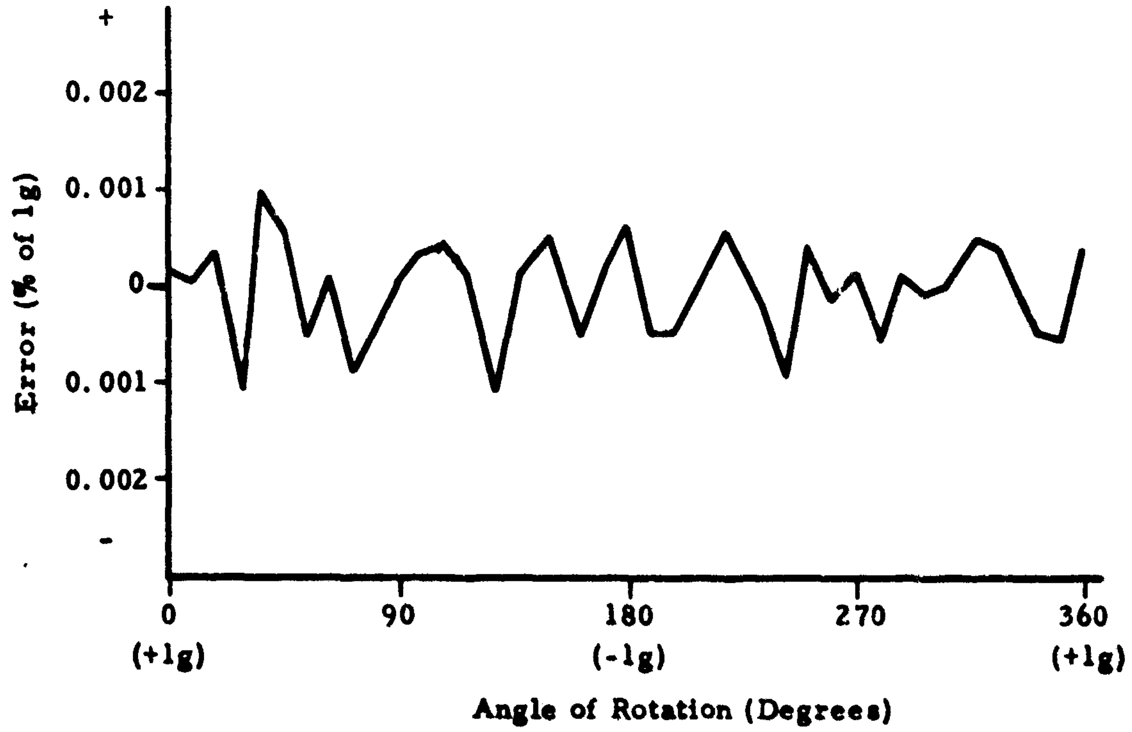


Fig.23 Zero to one g linearity

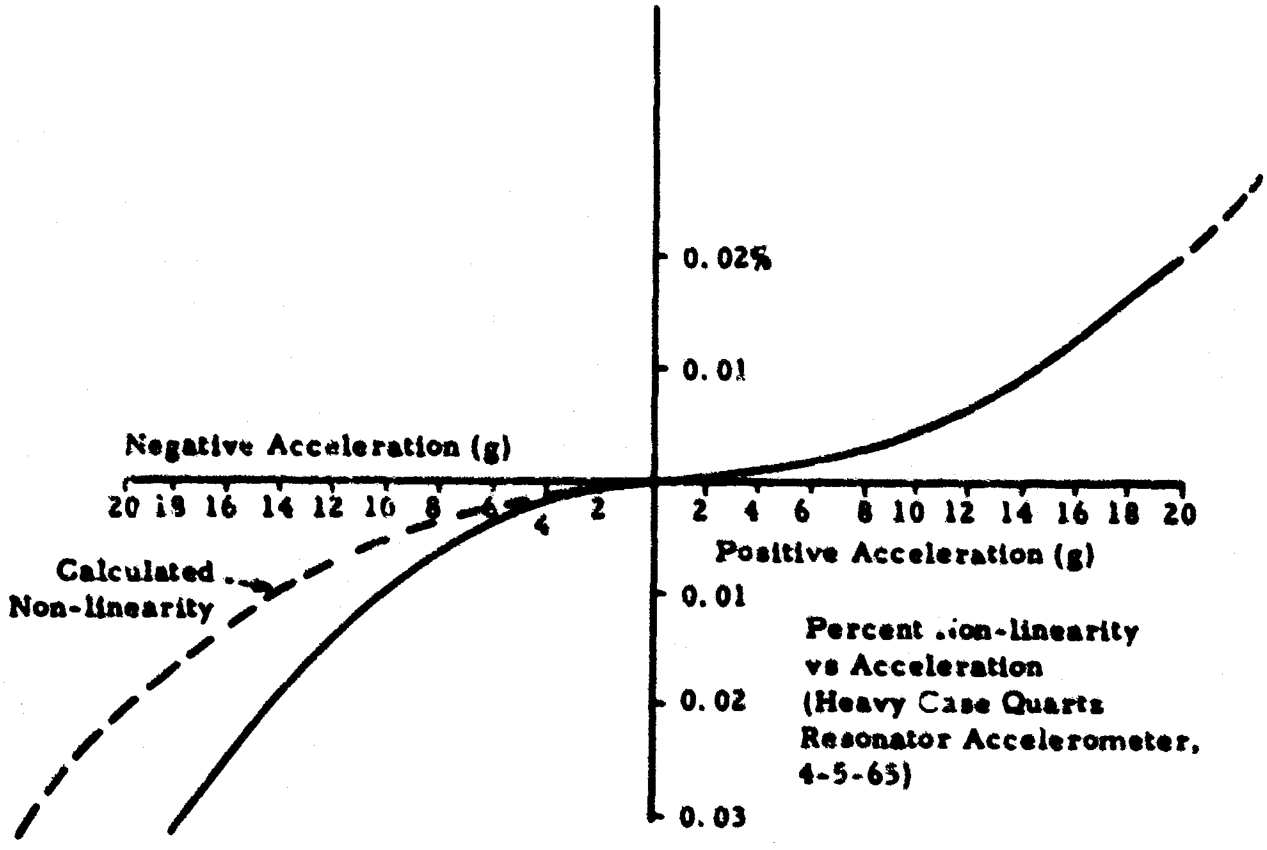


Fig.24 20 g accelerometer linearity

AN OPTICALLY PUMPED NUCLEAR MAGNETIC RESONANCE GYROSCOPE

by

James H. Simpson and I. A. Greenwood

General Precision Systems Inc.
Pleasantville, NY, USA

SUMMARY

The nuclear magnetic resonance (NMR) gyroscope described is a rate integrating single input axis instrument. It has no moving parts; many of its advantages, such as extreme acceleration tolerance and lack of reaction torques arise primarily from this feature. In addition, the nuclear gyro can accommodate large input rate and angle ranges.

Certain nuclides possess an intrinsic angular momentum and hence a nuclear magnetic dipole moment. The sensing element of the nuclear gyroscope, low density mercury vapor, contains a collection of such atomic nuclei. This collection of mercury nuclei behaves very much like many microscopically small but excellent gyroscopes having bar magnets on their spin axes and without mechanical interactions with their surroundings. The apparent changes in the dynamic behavior of the collection of nuclei measured by the surrounding instrument when it rotates with respect to inertial space provides the rotation sensing information. More particularly, changes of precessional motion of the nuclear magnetization in a magnetic field during rotation of the apparatus are detected.

The use of two different kinds of nuclei, ^{199}Hg and ^{201}Hg, eliminates the need for very accurate control or knowledge of the magnitude of the applied magnetic field. NMR signals with large signal-to-noise ratios are obtained with only weak (1.2 gauss) magnetic fields through the use of optical pumping for orientation of nuclei and through optical monitoring of the precession motion.

The instrumentation of a recently tested experimental model is described. In this model, phase comparisons between output signals from NMR controlled oscillators supply the rotation output information. Design factors affecting random drift rates and application ranges are discussed.

AN OPTICALLY PUMPED NUCLEAR MAGNETIC RESONANCE GYROSCOPE*

James H. Simpson and I.A. Greenwood

The optically pumped nuclear magnetic resonance gyroscope is a rotation sensing device that measures the integrated rate of rotation about a single axis. In this instrument, the rotation sensing element is a collection of oriented atomic nuclei. Aside from the thermal motion of the atoms there are no moving parts; therefore, the operation of the nuclear gyroscope should not be influenced by high accelerations and should create no reaction torques. Other desirable characteristics will be shown to be derived from this characteristic.

In this paper we shall state briefly the fundamentals of operation, examine one form of a laboratory instrumentation, and discuss the operation and application of the gyro.

Perhaps the basic motivation for the development of a nuclear gyroscope came from the simplified picture of certain atomic nuclei, which possess inherent angular momentum. These nuclei act in many ways as if they were microscopic spinning masses at the centers of atoms. The spinning motion requires no power input and the gimbaling is nearly perfect. Of course, since the nucleus is positively charged, it will also possess a magnetic moment. The ratio of the magnetic moment to the mechanical angular momentum is a characteristic of each nuclide that has angular momentum and is called the gyromagnetic ratio.

In order to see how rotation information can be obtained from a collection of oriented nuclei, consider the dynamic motion of such a collection in an applied constant magnetic field H_0, as illustrated in Figure 1. The net magnetization $\vec{M}$ is the resultant of all the individual magnetic moments. The general motion of the magnetization $\vec{M}$ in the magnetic field is one of precession about the direction of the magnetic field. The precessional angular frequency is called the Larmor frequency and, as measured in a non-rotating frame, is $\omega = \gamma H_0$. Rotation information sensed arises from changes in the phase of this precession when the observer's apparatus rotates about the direction of the magnetic field. In order to separate the shifts in phase due to rotation from those due to changes in the magnetic field, two kinds of nuclei may be used and hence two nuclear phases observed. Thus, knowledge or exact control of the magnetic field is not required[1,2].

A schematic drawing of the optically pumped NMR gyroscope is shown in Figure 2. The two nuclei with which the magnetic resonance is observed are the stable odd isotopes of mercury: ^{199}Hg and ^{201}Hg. The mercury is in the form of a monatomic vapor (about 10^{-4} mmHg pressure or $\sim 10^{12}$ atoms/c.c.) and is contained in a cylindrical cell about 1 cm in each dimension. The mercury absorption cell is made of high-purity synthetic fused silica. The frequencies at which the signals occur are 1 kHz for ^{199}Hg and 369 Hz for ^{201}Hg in the 1.2 gauss magnetic field that is used. The signals are observed as the outputs of NMR-controlled oscillators in which optical pumping is employed to enhance the net nuclear magnetization and also to monitor the precessional motion of the nuclear magnetization. Optical pumping refers to the exchange of angular momentum between the circularly-polarized pumping beam and the mercury atoms. By this process the fractional orientation of mercury nuclei is enhanced by a factor of about 10^9. The precessional motions of the magnetization cause the monitor beam to be modulated at the two frequencies. The signals on the light beam are converted to electrical signals, amplified, and fed-back to drive the magnetization in a steady precession. In Figure 3 we see a schematic diagram of the arrangement with which most of the test operations have been performed.

---

* Currently funded through Contract N00017-67-C-2419.

Here the two beams have been combined into one beam at 45^0 with respect to the magnetic field with which both pumping and readout are accomplished. The light from a single beam is divided into two beams to pump and readout two mercury cells, each of which contains both mercury isotopes. The two NMR-controlled oscillators are arranged with their constant magnetic fields pointing in opposite directions, so that rotations which add to the phases of the signals in one oscillator subtract from the phases in the other loop. The rotation information is extracted from the outputs of the NMR-controlled oscillators. This arrangement facilitates the phase comparison of the NMR signals and also tends to minimize the effects of light fluctuations. The NMR components are shown in Figure 4. The model shown has been subjected to test operation in the laboratory, as have several models that preceded it.

Now consider the fundamental experimental parameters and their influence on gyro performance. Fluctuations in any parameter which lead to a deviation in the frequency differences at the output of the spin generators will contribute to uncontrolled drift in the output of the gyro. The various sources of random drift are listed in Table I. The list is divided into two parts. The first part contains those sources for which there is no correlation between the two spin generator loops. The second part contains those effects which need not be minimized on an absolute basis but only on a differential basis.

Both the noise in the spin generator loop and the fundamental magnetization phenomena lead to random fluctuations in the phase of the oscillations of the spin generators. The primary source of the noise in the spin generators is the noise arising from the photon beam and its detection. Its effect is minimized by achieving large signal-to-noise ratios in the magnetic resonance cells. Current values well in excess of 40 dB put this contribution below several thousandths of a degree per hour.

The magnetization noise arises from the statistical nature of the collection of nuclear magnetic moments which form the nuclear magnetization. The effects of these fluctuations are minimized by attaining long NMR damping times and a very high percentage orientation. Estimates of this contribution[2] show this to be of the same order of magnitude as the shot noise.

The phase stability of the electronic components, particularly the amplifier in the spin generator loop, determines the degree to which the output of the oscillator follows the ideal precessional motion of nuclear magnetizations. Through the use of large amounts of feedback, the desired stability can be achieved.

A shift in the effective direction of the H_0 field, due to, let us say, a change in the ambient field, has two effects on the gyro output. The first is that it shifts the direction of the sensitive axis of the gyro so that it no longer coincides exactly with the externally labelled axis. A small amount of magnetic shielding can reduce this shift to the point of being not important. The second effect may be to introduce a phase shift through a change in the effective angle between the H_1 field and the light beam. A preferred relative orientation of the H_1 field coil and the light beams reduces this effect and makes it negligible.

The observed Larmor frequency, that at which the magnitude of the resonance reaches a maximum and the phase shift achieves a medial value, is a function in an optically pumped experiment, primarily of the intrinsic gyromagnetic ratio and the magnetic field. It is also weakly dependent on intensity, polarization, and spectral distribution of the incident beams of light[3]. Through care in equalizing the light in the two cells, the effects of changes in light intensity may be balanced in the two spin generators. This is accomplished in the apparatus that is shown by deriving the two beams of light from a single beam, through the use of the reflection polarizers.

The parameters presented in Table I represent the major problem areas that have been identified and studied. In each case a satisfactory solution has been found to eliminate or minimize its effect on the performance of the gyro. The more recent work has been concerned with the development of procedures and techniques which will result in the reduction of the laboratory instrumentation to one of practical size and power requirements. The design of the latest models under construction shows these changes.

Some of the desirable characteristics of the optically pumped NMR gyroscope are summarized briefly in Table II. Most of the characteristics are derived from the fact that there are no moving parts. The lack of moving parts should lead to a long operating lifetime, relative ease of construction and maintainance, and low cost. Since all components of the gyro are firmly fixed in the structure, the instrument should be capable of operating and providing accurate readout during periods of very high accelerations. The absence of a large rotating mass also eliminates the possibility of a reaction torque being exerted. Since there are no gimbals forced to follow rapid motions, the gyro can accommodate the wide range of input rates and angles associated with strapdown applications.

The use of any unconventional gyro in a guidance system should not be thought of as simply replacement of a conventional mechanical gyro. In order to fully realize all the unique advantages of a new device, the system integration should be thought through completely in terms of the component input and output characteristics; for example, in the case of the nuclear gyro, the full information stored in the phase of the precessing nuclear magnetization should be utilized.

At the present time, the studies of the nuclear gyroscope have demonstrated the validity of the concept of using two different nuclides in the same magnetic field to eliminate exact control of the magnetic field. Most of the work on reduction toward a useful instrument has been accomplished, in that critical problem areas have been identified and practical solutions have been found. In the next phase of investigation, instrument evaluation, the range of performance characteristics of the instrumentation will be determined.

Through additional investigation of the fundamental phenomena associated with optically pumped nuclear magnetic resonance, improvements in basic limiting parameters such as signal-to-noise ratios, magnitudes of lamp shifts, and phase stability, would lead to a second group of nuclear gyroscopes with greater accuracy, smaller size, and reduced power consumption.

REFERENCES

1. Simpson, J.H. Memorial d'Artillerie française, Vol.39, 3^e fasc., 1965, p.463.

2. Simpson, J.H. *Experimental Model Nuclear Gyroscope*. AD 431-727.

3 Cohen-Tannoudji, C. Ann. Phys., Vol.7, 1962, pp.423 and 469.

TABLE I

Sources of Random Drift in the Nuclear Gyro

1. Noise in spin generator loop.
2. Magnetization noise.
3. Phase stability of electronic components.
4. Changes in effective direction of the constant magnetic field.
5. Light induced shifts in Larmor frequency.

TABLE II

Characteristics of the Nuclear Gyro

1. Capable of operating and providing accurate readout during periods of high acceleration.
2. Wide range of input rates and angles.
3. No reaction torque.
4. Long operating lifetime.
5. Relative ease of construction and maintainance, low cost.

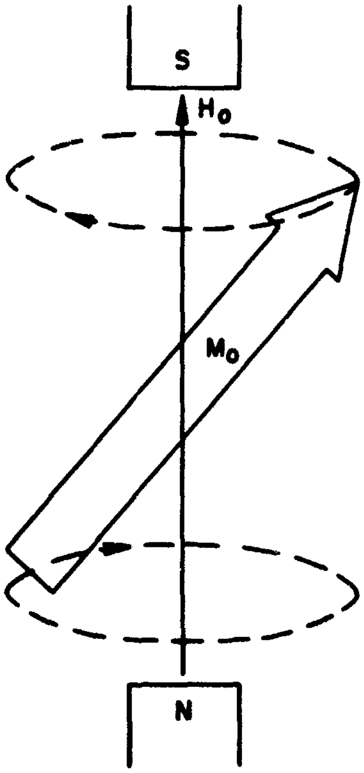


Fig. 1 Precession of nuclear magnetization about magnetic field

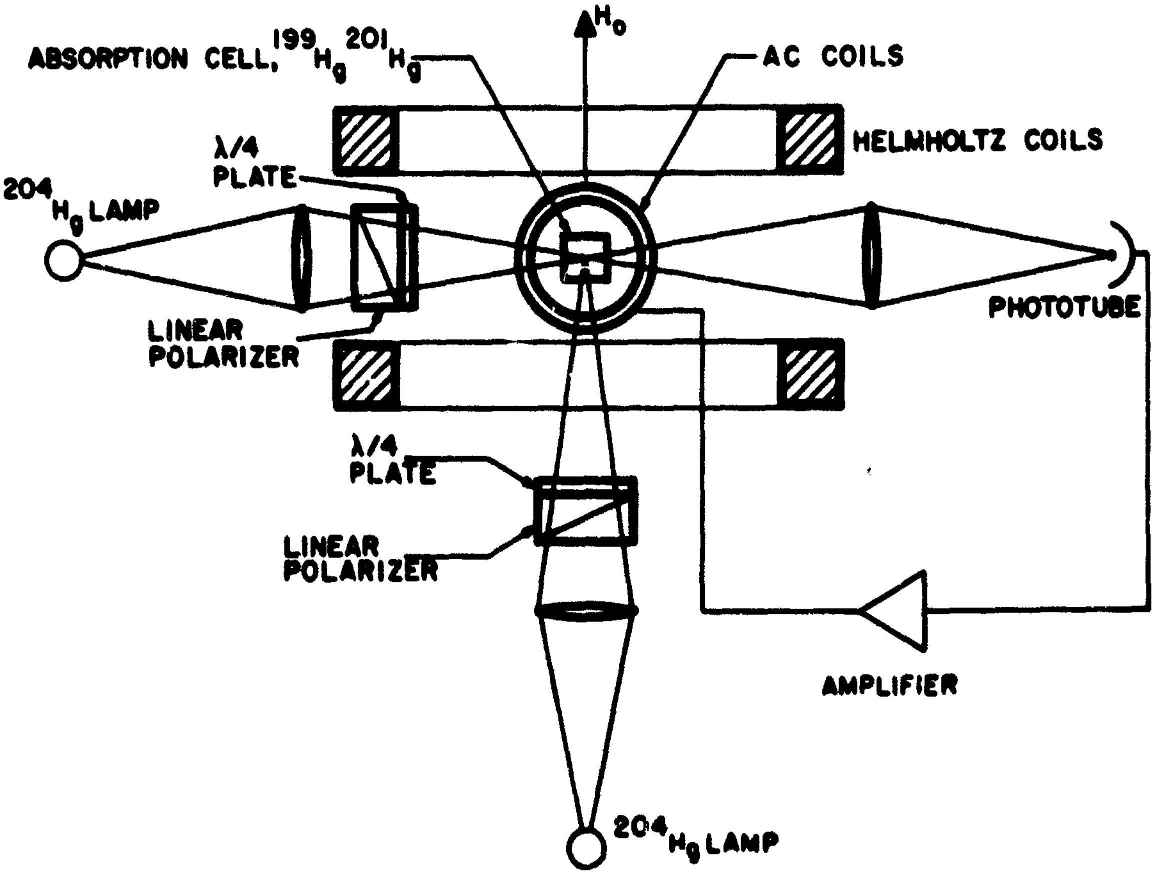


Fig. 2 NMR-controlled oscillator with optical pumping

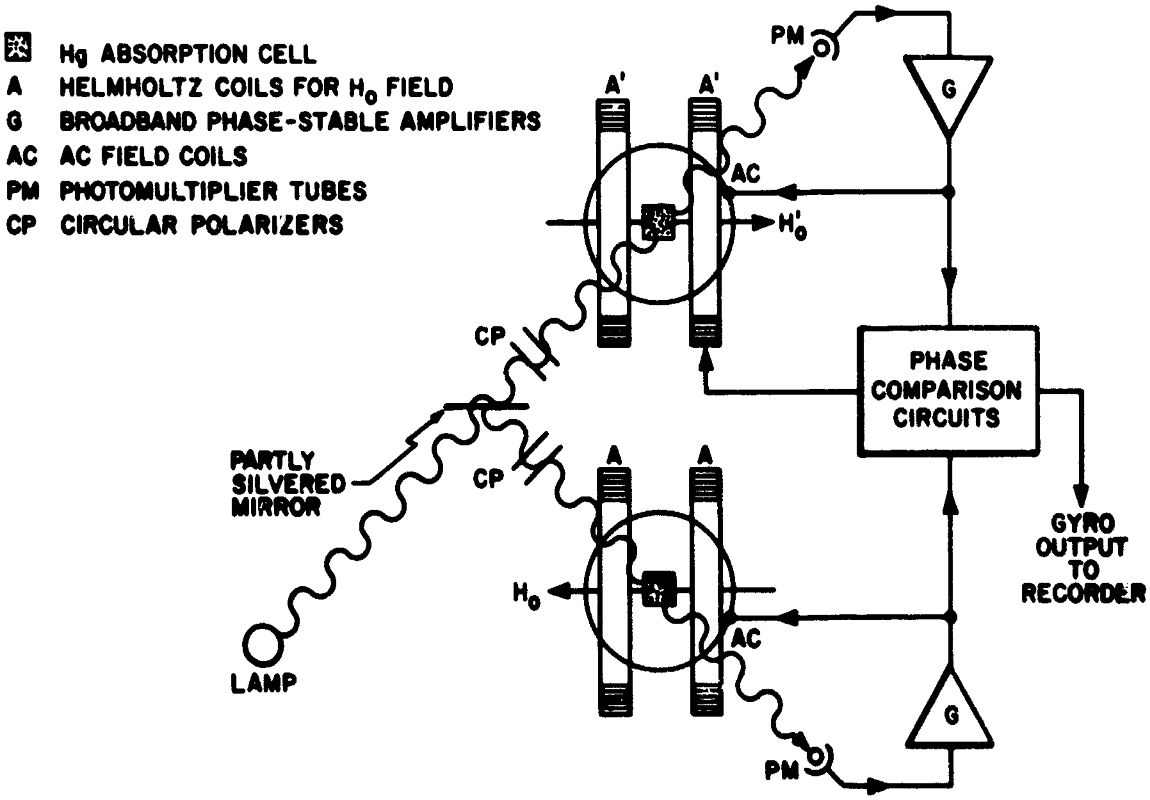


Fig.3 Optically pumped NMR gyroscope

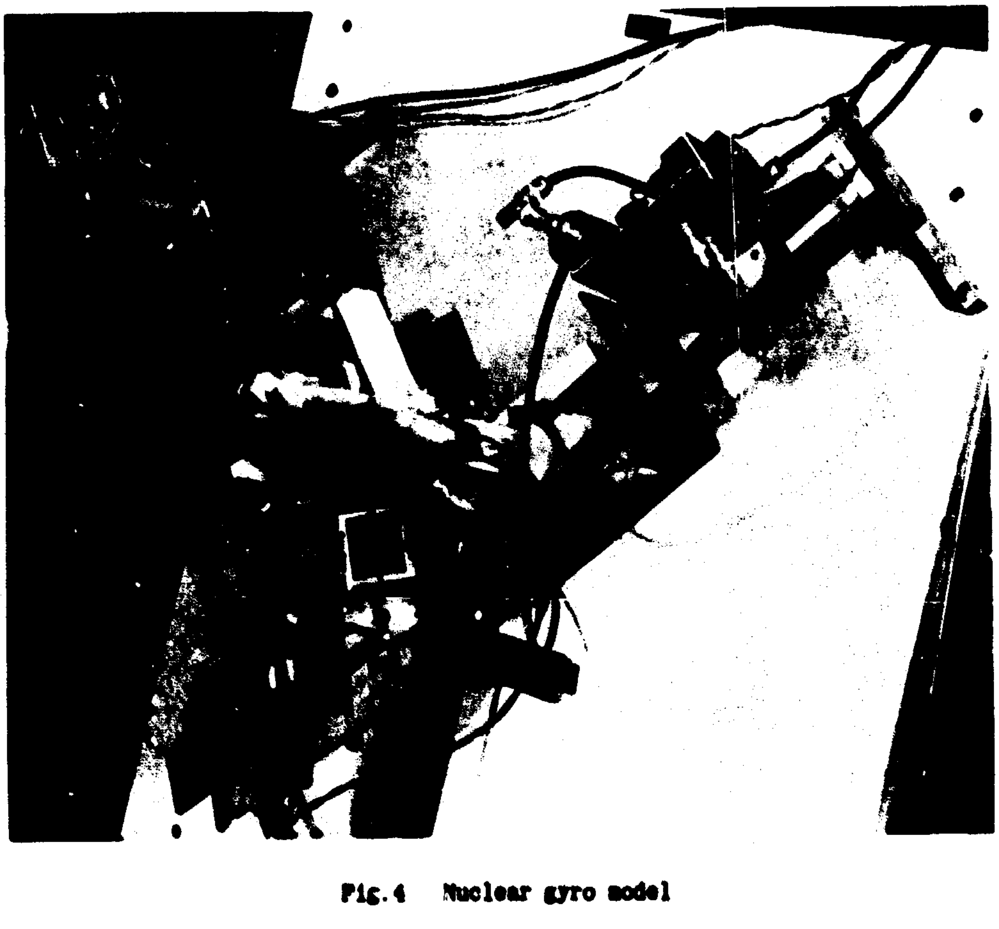


Fig.4 Nuclear gyro model

GYROMETRES LASER ET APPLICATIONS

par

J.M.Catherin et B.Dessus

Centre de Recherches de la Compagnie Générale d'Electricité,
54, rue La Boëtie, Paris, 8e, France

GYROMETRES LASER ET APPLICATIONS

J.M.Catherin et B.Dessus

1. INTRODUCTION

Le principe du gyromètre à laser est maintenant bien connu des spécialistes de navigation et de guidage par inertie. Néanmoins, nous le rappellerons rapidement pour faciliter la compréhension de ce qui suit.

Un gyromètre à laser est constitué d'un tube dans lequel on produit un plasma susceptible d'amplifier de la lumière, ce tube étant placé dans une cavité optique fermée sur elle-même.

La forme la plus simple est un triangle constitué par des miroirs ou des prismes à réflexion totale. Nous avons, alors, un dispositif composé d'un amplificateur couplé à un résonnateur et soumis à une rétroaction presque totale puisque 97 à 99% de l'énergie émise par le tube y rentre de nouveau.

Ce système et donc susceptible d'osciller, si le gain du tube est supérieur aux pertes de la cavité.

Le laser émet alors deux ondes progressives qui parcourent la cavité en sens contraire. La fréquence d'oscillation est celle qui aura les pertes minimales: c'est à dire, celle dont le déphasage au cours d'une révolution est un multiple entier de 2π radians, ou encore, celle dont la longueur d'onde est sous multiple de la longueur de trajet optique L .

A très peu de chose près, c'est la cavité qui fixe la fréquence.

- Si le gyromètre à laser est en repos ou en translation dans un espace absolu, le trajet des deux ondes progressives est le même, les fréquences sont les mêmes.
- Si le gyromètre à laser est en rotation dans ce même espace, une onde voit alors son trajet allonger, l'autre voit son trajet diminuer, les deux fréquences d'oscillation sont différentes et, si l'on rend colinéaires les rayons transmis par un miroir, on obtient sur un détecteur quadratique un battement dont la fréquence est la différence des fréquences des deux ondes progressives.

La fréquence de battement est donnée par

$$\Delta\nu = \nu_2 - \nu_1 = \frac{4S}{\lambda P}\Omega .$$

ν_1, ν_2	fréquence de chacune des deux ondes progressives
S	surface délimitée par le trajet de ces ondes
P	périmètre du trajet optique
λ	longueur d'onde de travail.

De même que la fréquence est liée à la vitesse angulaire, la phase est liée à l'angle, et une mesure d'angle peut se faire par simple comptage du nombre de périodes de battement délivré par le gyromètre à laser au cours du balayage de cet angle, ce nombre étant indépendant de la vitesse de parcours:

$$N\alpha_2 - \alpha_1 = \frac{4S}{\lambda P}(\alpha_2 - \alpha_1) \; .$$

En toute rigueur, l'explication correcte du phénomène s'obtient à partir des principes de la relativité généralisée: la vitesse de la lumière n'est pas constante dans un référentiel tournant[11].

Avant de poursuivre, citons quelques ordres de grandeur en prenant par exemple le gyromètre à laser Hélium Néon fonctionnant sur une raie de longueur d'onde 6328 Å:

fréquence: $\nu_0 \simeq 4.5 \times 10^{14}$ Hz

largeur Doppler: $\Delta\nu_D \simeq 10^9$ Hz

surtension de la cavité: quelques 10^8 soit une

largeur de bande $\Delta\nu_c$ quelques 10^6 Hz

largeur de raie de l'oscillation $< 10^{-3}$ à 10^{-4} Hz .

Enfin, pour un gyromètre triangulaire de 15 cm de côté on a sensiblement

$$\Delta\nu \simeq 1 \times \Omega \; ,$$

avec $\Delta\nu$ en Hz et Ω en degrés par heure, soit

$$N \simeq 1 \times \Delta\alpha \; ,$$

où N est le nombre de périodes comptées et $\Delta\alpha$ le déplacement angulaire en secondes d'arc.

2. AVANTAGES DES GYROMETRES A LASER

Le gyromètre à laser est un dispositif statique: c'est la première caractéristique le différenciant des gyromètres conventionnels. Il est inutile d'insister sur les avantages que présente cette propriété.

Le gyromètre à laser a une dynamique très étendue avec une excellente linéarité: entre quelques centaines de degré/heure et 10^7 °/h (soit 500 t/mn environ) les écarts de linéarité sont inférieures à 10^{-6}.

Le gyromètre à laser fournit ses mesures directement sous forme digitale, fréquence pour les vitesses, nombre d'impulsions pour les angles. Pour un gyromètre laser triangulaire de 15 cm de côté, on obtient aisément une résolution de ¼ de seconde d'arc.

Le traitement des mesures par calcul est donc particulièrement aisé. On évite en particulier le problème de l'exploitation des faibles signaux analogiques. Le gyromètre à laser est faiblement sensible à la température et, par principe, insensible aux accélérations et vibrations. La limite pratique de cette insensibilité est donnée par la rigidité de la structure optique, cette rigidité pouvant être portée à un degré très élevé.

Enfin, la durée de vie est celle du tube laser, c'est à dire plusieurs milliers d'heures.

3. LES DIFFICULTES DU GYROMETRE A LASER ET LES SOLUTIONS POSSIBLES

Prises par ordre d'importance au niveau des principes, ces difficultés sont les suivantes:

- Phénomène de couplage entre les ondes optiques
- Non réciprocité du trajet optique dans le milieu amplificateur
- Conservation de l'alignement du trajet optique malgré l'influence de la température, des vibrations et des contraintes mécaniques.

Reprenons dans l'ordre inverse chacune de ces difficultés en explicitant le problème et en citant les remèdes déjà utilisés ou à l'étude

La précision nécessaire d'alignement des miroirs ou prismes d'un gyromètre à laser est de l'ordre de quelques secondes d'arc. On voit tout de suite le rôle que peuvent jouer la température et les perturbations mécaniques sur une cavité optique mal conçue. De plus, nous verrons par la suite qu'il n'est pas suffisant de rester dans les conditions d'oscillation du laser, mais qu'il faut à cette oscillation une grande stabilité de puissance et de fréquence si l'on veut atteindre une certaine précision. Il existe actuellement plusieurs solutions permettant d'obtenir des structures optiques très rigides et stables en température. Elles ont en commun de supprimer les réglages d'alignement des miroirs ou des prismes, ceux-ci étant directement collés sur une pièce massive en quartz ou en granit. Il est néanmoins très utile, dans certaines applications, de conserver un élément optique monté sur une céramique piézo électrique et d'avoir ainsi la possibilité de stabiliser de façon très précise la longueur de cavité et, par conséquent, la fréquence d'oscillation. Ces solutions peuvent donner au gyromètre à laser une nette supériorité pour l'application aux engins à fortes accélérations.

Dans la technique que nous utilisons, nous disposons d'un tel élément piézo-électrique. La limite de la variation de température admissible sur ce gyromètre à laser sera déterminée par la correction maximale de longueur que l'on peut obtenir de l'élément piézo, soit une dizaine de microns. La variation correspondante de température est d'une dizaine de degrés centigrades avec un support en granit, quelques dizaines avec un support en quartz.

Il est possible que dans une étape ultérieure, on soit amené à stabiliser la température pour pousser le gyromètre à sa précision ultime, lorsque les erreurs du second ordre dues aux variations de température ne seront plus négligeables.

Nous avons cité en second lieu, une difficulté moins connue des gyromètres à laser: la variation non réciproque du trajet optique dans le milieu amplificateur

En effet, aux débuts des études sur le gyromètre à laser, les efforts ont porté essentiellement sur les problèmes de couplage et sur la recherche de diverses méthodes de faux zéro permettant de contourner la difficulté. Par la suite, on s'est intéressé à ces phénomènes de non réciprocité car leur influence sur les erreurs du gyromètre à laser n'étaient plus négligeables. A mesure que les gyromètres à laser gagneront en précision, ces phénomènes prendront une importance de plus en plus fondamentale.

La relation $\Delta\nu = (4S/\lambda P)\Omega$ n'est exacte que pour une cavité passive, car le milieu amplificateur peut introduire une différence de trajet optique qui vient se composer avec celle introduite par une rotation.

En effet, nous avons dit que le gain de ce milieu n'est pas constant avec la fréquence. La caractéristique est sensiblement celle de l'élargissement Doppler de la raie d'émission, c'est à dire une Gaussienne. Dans le cas des filtres électriques, Bode a montré qu'une variation de la transmittance avec la fréquence s'accompagnait d'une variation de phase. De même, les relations de Kramers Kronig montrent dans le cas des diélectriques qu'une variation d'absorption avec la fréquence s'accompagne d'une variation d'indice

Le phénomène est le même en émission stimulée, donc dans le cas d'un gain (Fig. 1).

Cette variation d'indice, qui est très faible, va engendrer une variation de longueur de trajet optique, proportionnelle à la longueur du milieu amplificateur traversé.

La formule précédente devient alors

$$\Delta\nu = \nu_2 - \nu_1 = \frac{4S}{\lambda P}\Omega + \frac{CL_a[n(\nu_2) - n(\nu_1)]}{\lambda P} .$$

où L_a est la longueur du milieu amplificateur traversé et $n(\nu)$ est l'indice du milieu en fonction de la fréquence optique.

Cette fonction est tabulée et s'appelle fonction de dispersion d'un plasma. Son expression analytique est assez compliquée et nous en prendrons un développement limité valable au voisinage du centre de la raie (ce qui correspond effectivement à nos conditions de travail étant donné que nous employons un mélange moitié moitié de Ne_{20} et de Ne_{22} destiné à éviter la compétition entre les deux ondes progressives dans cette zone de fréquence).

Un calcul simplifié destiné seulement à donner des ordres de grandeur donne

$$n(\nu) = 1 + 0,5 \times 10^{-8} \frac{\nu - \nu_0}{\Delta\nu_D} \left[1 + \frac{\Delta G}{G_0}\right] \left[1 - \left(\frac{\nu - \nu_0}{\Delta\nu_D}\right)^2\right]$$

Cette formule a été obtenue en prenant un gain de 4% pour le laser. Les bases de ce calcul peuvent se trouver dans la Référence 3. En portant cette formule dans la précédente et en supposant que ν_1 est maintenue égal à ν_c par l'asservissement de fréquence on obtient

$$\Delta\nu = \nu_2 - \nu_1 = \frac{4S}{\lambda P}\Omega \left\{ \frac{1}{1 - \frac{CL_a(0,5 \times 10^{-8})}{\Delta\nu_D \lambda P} \left(1 + \frac{\Delta G}{G_0}\right) \left[1 \left(\frac{\nu_1 - \nu_2}{\Delta\nu_D}\right)^2\right]} \right\} ;$$

la signification de ν_0 et $\Delta\nu_D$ a été donnée plus haut, le facteur $1 + \Delta G/G_0$ indique l'influence d'une variation de gain ΔG autour d'une valeur G_0 de ce gain.

Donnons une application numérique:

avec C , vitesse de la lumière = 3×10^8m/s

$\lambda = 0,6 \times 10^{-6}$m

$L_a = 0,15$ m

$P = 0,45$ m

$\Delta\nu_D = 10^9$ Hz

$\frac{4S}{\lambda P} = 2 \times 10^5$

on a

$$\Delta\nu \approx \frac{2 \times 10^5}{1 - 10^{-3}(1 + \Delta G/G_0)\left[1 - \left(\frac{\Delta\nu}{10^9}\right)^2\right]} \Omega ;$$

avec Ω en rad/s soit

$$\frac{-1}{1 - 10^{-3}(1 + \Delta G/G_0)\left[1 \left(\frac{\Omega}{10^9}\right)^2\right]}\,\Omega$$

si l'on exprime Ω en degrés par heure. L'écart de linéarité ϵ est alors

$$\epsilon = 10^{-3}\,\frac{\Omega^2}{10^{18}}\ .$$

Ce résultat montre que même pour des vitesses aussi élevées que 10^7 °/h l'erreur de linéarité $\epsilon = 10^{-7}$ sera tres faible. Les variations de gain auront une influence beaucoup plus importante en modifiant le facteur d'échelle.

A cela s'ajoutent les effets de saturation que présentent tous les oscillateurs. Dans le cas des lasers employés pour les gyromètres, la raie d'émission est inhomogène, c'est à dire que la variation de gain engendrée par la saturation reste localisée autour de la fréquence d'oscillation au lieu de se répartir sur l'ensemble de la raie.

Cela signifie que les distorsions dans la courbe de gain, donc dans la courbe d'indice, sont relativement marquées. Le trajet optique de chaque onde dépend donc du degré de saturation qu'elle provoque dans le milieu amplificateur, donc de la puissance, ou encore de la différence gain moins pertes de la cavité pour ce sens de propagation. Les pertes peuvent, en effet, être légèrement différentes pour les deux sens de parcours à cause des phénomènes de diffusion.

La différence de trajet ainsi obtenue se manifeste même à vitesse de rotation nulle, c'est à dire lorsque $\nu_1 = \nu_2$. Pour ces problèmes voir les Références 4-6 et 8.

Ces phénomènes montrent qu'il est nécessaire de stabiliser au mieux la fréquence d'oscillation, le gain du laser et les pertes de la cavité.

La stabilisation de fréquence, peut être obtenue, soit en comparant la fréquence optique du gyromètre à celle d'un laser stabilisé annexe par hétérodynage optique, soit en stabilisant le gyromètre sur lui-même, c'est à dire au sommet de la courbe de gain du milieu amplificateur.

Les deux solutions ne présentent pas de grandes difficultés et les résultats sont satisfaisants.

La stabilisation du gain implique l'emploi d'alimentations stables et la stabilisation des pertes est une affaire de rigidité et de propreté de la cavité optique. Mais, cela n'est pas suffisant pour certaines applications. Il est alors nécessaire de mesurer les variations résiduelles et d'introduire une correction appropriée.

On a calculé une formule théorique reliant divers paramètres importants; elle a la forme

$$F_2 - F_1 = (\nu_2 - \nu_1)\left[A - B\,\frac{(I_1 + I_2)}{2}\right] + C(I_2 - I_1)\ .$$

avec $\nu_2 - \nu_1$ fréquence mesurée, $F_1 - F_2$ fréquence corrigée, I_1 et I_2 puissance des deux ondes. A , B and C sont des paramètres dépendant principalement du gain de la fréquence et de diverses constantes physiques du milieu amplificateur, mais ne dépendent pas des pertes de la cavité qu'il n'est donc pas nécessaire de connaître. La fréquence étant fixée par ailleurs, on peut mesurer des quantités proportionnelles à I_1 et I_2, seul le gain reste actuellement très difficile, sinon impossible, à mesurer avec précision au cours du fonctionnement du gyromètre à laser. Cela n'est pas très grave dans la mesure où les variations de puissance sont dues beaucoup plus aux variations de pertes qu'aux variations de gain.

Cette formule va être vérifiée expérimentalement et dans la mesure où elle fournira des valeurs invariantes de A, B et C pour un laser donné, on pourra ensuite corriger la mesure donnée par le gyromètre par un calcul fait en temps réel. Il est, en effet, de loin préférable d'effectuer une correction par calcul que de fixer I_1 et I_2 à des valeurs de consigne par un asservissement. Celui-ci apporte en effet plus d'inconvénients que d'avantages.

Nous avons placé en dernier lieu l'étude du phénomène de couplage car le choix des solutions à ce problème dépend en partie des considérations qui ont précédé dans l'étude de la non réciprocité des trajets optiques.

On a observé, depuis longtemps, qu'en dessous d'une certaine valeur de la vitesse de rotation, le battement disparaissait totalement, ce qui signifie que les deux ondes émises par le laser sont à même fréquence, bien que les trajets optiques vus par chacune d'elles soient légèrement différents. Ce phénomène est imputable en premier lieu aux imperfections des éléments optiques placés sur le trajet des deux ondes et en second lieu aux caractéristiques du milieu amplificateur lui-même.

En efft, les irrégularités résiduelles des surfaces optiques provoquent la rétrodiffusion d'une partie de l'énergie incidente dans l'angle solide délimitant l'onde de sens contraire.

La situation est celle de deux oscillateurs faiblement couplés dont le désaccord relatif décroît. En dessous d'une certaine valeur de ce désaccord, les deux oscillateurs se synchronisent en fréquence et en phase. Dans un laser en anneau, au lieu d'avoir deux ondes progressives indépendantes, on a dans ce cas une onde stationnaire. Bien que de valeur relative très faible (quelques centaines de hertz de seuil pour 5×10^{14} hertz de fréquence d'oscillation), ce phénomène introduit un seuil de mesure de l'ordre de quelques centaines de degrés par heure qui rend le gyromètre à laser impropre à toute application de classe inertielle sans emploi d'une méthode de faux zéro.

La réduction du couplage est essentiellement une question de qualité et de propreté des divers éléments optiques et de simplicité de la structure. Tout élément ajouté sur le trajet des rayons lumineux est une source de rétrodiffusion, donc de couplage. De toute façon, quelles que soient les précautions prises, on abaisse difficilement le seuil de couplage en dessous d'une centaine de degré/heure, ce qui est de très loin insuffisant pour éviter l'emploi d'un faux zéro.

On a cherché de nombreuses solutions qui peuvent se ramener à deux types.

Tout d'abord des moyens visant l'obtention au repos d'une fréquence de battement aussi stable que possible. Ces moyens sont équivalents à une vitesse de rotation constante ajoutée à la vitesse mesurée.

Parmi ces moyens on peut citer l'effet Zeemann, l'effet Faraday et l'effet Fizeau[2,5,6,8].

L'explication détaillée de ces différents effets sortirait nettement du cadre de cet exposé. Disons, simplement, que dans les deux premiers la fréquence de battement est proportionnelle à un champ magnétique. Ces deux effets ne s'obtiennent qu'au prix d'une complication notable de la cavité optique et la stabilité du faux zéro est celle du champ magnétique.

L'effet Langmuir qui accompagne tout passage d'un courant continu dans un gaz provoque un effet Fizeau. Le courant électrique existe déjà le plus souvent car c'est le moyen le plus commode pour produire le plasma amplificateur du laser. Il est donc tentant d'utiliser cet effet comme faux zéro. Sa stabilité, qui dépend du courant et de la pression, est meilleure que 10^{-4}.

Ces procédés conviennent parfaitement pour mesurer des vitesses de rotation avec pour seule condition que la somme algébrique de la vitesse fictive correspondant au faux zéro et de la vitesse à mesurer ne soit pas inférieure en valeur absolue au seuil de couplage. Cela impose, soit de ne mesurer que des vitesses de rotation dans un seul sens, soit que la plus grande vitesse de rotation à mesurer soit inférieure à la valeur de la vitesse de faux zéro. Comme la précision du faux zéro est définie par une erreur relative et que les erreurs dues au milieu amplificateur sont aussi pour l'essentiel des erreurs relatives, de grandes valeurs de ce faux zéro peuvent conduire à une erreur absolue prohibitive. On peut aussi avec ce procédé mesurer des angles avec une bonne précision à condition que la mesure soit faite dans un temps court. Sinon, l'erreur angulaire étant égale alors au produit de l'erreur de vitesse et du temps de mesure, on peut aboutir à des valeurs importantes de cette erreur pour des temps de mesure importants. C'est le problème des mesures faites par intégration.

Le second procédé consiste à obtenir ce faux zéro à partir d'une vitesse de rotation mécanique dont l'intégrale est cette fois directement mesurable puisque c'est un angle. Remarquons tout de suite que dans ce cas, si l'on utilise le gyromètre en mesure d'angle, la vitesse de rotation mécanique produisant le faux zéro n'a pas besoin d'être constante.

Citons tout d'abord le procédé mis au point et utilisé par Honeywell qui consiste à faire vibrer l'ensemble de la structure au tour d'une position fixe. Ici, pas de problème de mesure du faux zéro puisque l'angle moyen de rotation est nul. La mesure est obtenue en effectuant la soustraction du nombre d'impulsions délivré par le gyromètre au cours de sa rotation dans un sens et dans l'autre. A chaque demi-période, le gyromètre passe donc dans la zone de couplage.

Il existe ensuite deux autres procédés, l'un consistant à faire tourner le gyromètre autour de son axe sensible et à obtenir la mesure en effectuant la différence entre l'angle mesuré par le gyromètre et l'angle décrit par le gyromètre par rapport au support. Un procédé similaire consiste à faire tourner une lame transparente traversée par les rayons lumineux du gyromètre. A cause de l'effet Fizeau, le trajet optique parcouru par les deux ondes est différent.

La rotation de la lame est équivalente à une rotation fictive du gyromètre, et il existe un rapport bien défini entre les deux ne dépendant que des dimensions géométriques.

Une solution intéressante consiste à faire tourner réellement ou fictivement le gyromètre à laser à vitesse sensiblement constante par rapport à un référentiel de Galilée, c'est-à-dire à fréquence de battement constante. Il suffit d'asservir la vitesse de rotation du gyromètre par rapport à son support de façon à maintenir la fréquence de battement à une valeur de consigne.

Cette méthode a l'avantage de résoudre le problème de dynamique de mesure posé par l'existence de la zone de couplage, le support peut être animé d'une vitesse absolument quelconque. D'autre part, la fréquence de battement est fixée à une valeur suffisamment élevée pour se situer en dehors de la zone de non linéarité et de fluctuations dues au couplage, mais assez basse pour que les variations de facteur d'échelle engendrées par les variations de gain ou de puissance ne produisent qu'une faible erreur absolue. La fréquence choisie serait de l'ordre de quelques kilohertz.

On perd évidemment l'avantage du système purement statique, mais il faut remarquer que les performances requises de l'asservissement de vitesse sont très faibles, étant donné qu'il n'y a pas d'accélérations angulaires à compenser (donc peu d'énergie à dépenser) et qu'une erreur de vitesse de 20 ou 30% n'aurait pratiquement pas d'influence sur la précision du gyromètre.

4. DIVERSES APPLICATIONS

Suivant le type d'application recherché, le gyromètre à laser pourra être réalisé sous des formes et utilisé dans des systèmes relativement différents.

Dans des problèmes tels que la mesure d'angle dans un temps bref, le chercheur de Nord, la stabilisation d'engins, dans lesquels la notion de dérive a peu d'importance, ou bien il est inutile d'introduire un faux zéro, ou bien celui-ci n'a pas besoin de stabilité à long terme. On pourra utiliser, par exemple, l'effet Fizeau de la décharge.

Dans les problèmes de conservateurs de cap ou de navigation par inertie, la notion de dérive est au contraire essentielle et l'on aura recours aux moyens cités plus haut pour avoir un faux zéro stable et corriger les dérives dues au laser lui-même.

5. APPLICATION AU CHERCHEUR DE NORD

Avant de conclure, nous décrirons rapidement une application particulière du gyromètre à laser: le chercheur de Nord. Le problème est bien connu; il consiste à déterminer la direction de l'axe des pôles par une mesure de la rotation terrestre. Les propriétés de linéarité et de faible temps de réponse du gyromètre à laser permettent d'adopter une méthode dynamique de mesure éliminant l'effet de la dérive.

On fait tourner le gyromètre à laser autour de son axe sensible. Il donne alors une fréquence de battement qui varie proportionnellement au produit scalaire $\vec{n}.\vec{\Omega}$, où $\vec{n}$ est le vecteur unitaire porté par la normale au gyromètre à laser et $\vec{\Omega}$ le vecteur rotation terrestre.

Ce signal est donc une fonction sinusoïdale de l'angle de rotation imposé.

$$f = k\cos(\alpha - \phi) ,$$

avec ϕ l'angle compris entre la direction repère et la direction du Nord.

Si on effectue en temps réel le calcul d'un point de la fonction de corrélation du signal obtenu et d'une fonction sinusoïdale de l'angle α engendré par un codeur angulaire, on obtient

$$C = \frac{1}{2n\pi}\int_0^{2n\pi} k\cos(\alpha - \phi)\sin\alpha\, d\alpha = k\frac{\sin\phi}{2} \simeq k\frac{\phi}{2}$$

si ϕ est petit.

Comme on connaît k on peut déduire l'angle ϕ repérant la direction du Nord.

Cette méthode donne un rapport signal/bruit croissant proportionnellement au temps de mesure et permet donc d'éliminer assez rapidement les erreurs aléatoires.

On peut espérer ainsi obtenir des précisions convenables dans des temps nettement plus courts que ceux nécessaires à une mesure faite avec un gyro mécanique.

6. PERSPECTIVES

On peut considérer les limites ultimes des performances des gyromètres laser sous deux aspects différents.

Les limites purement théoriques ne dépendent que de la largeur de raie de l'oscillation laser. Cette largeur est tellement faible, qu'elle n'a pas encore pu être mesurée et son calcul exact nécessiterait une théorie quantique complète du laser en anneau. La théorie quantique actuelle n'est que partielle.

Néanmoins, on peut situer cette limite en dessous de 10^{-6} à 10^{-7} °/h (Ref.10).

En fait, les limitations pratiques viendront des possibilités de stabilisation ou de mesure des variations non réciproques du trajet optique dans le plasma amplificateur. Il est impossible de fixer actuellement des limites mais on peut espérer avoir dans un proche avenir 10^{-2} voire 10^{-3} °/h.

REFERENCES

1. Adler, R. — *A Study of Locking Phenomena in Oscillators.* Proceedings, Institute of Radio Engineers, June 1946, pp.351-357.

2. Macek, W.M.; Davis, D.T.M. — Journal of Applied Physics, Letters, Vol.2, 1963, p.67.

3. Aronowitz, F. — *Theory of Travelling Wave Optical Maser.* Physical Review, Vol.139, 1965, pp.A 635-6,A 646.

4. Lee, P.H.; Atwood, J.G. — *Measurement of Saturation Induced Optical Non-Reciprocity in a Ring Laser Plasma.* Institute of Electrical and Electronic Engineers, Journal of Quantum Electronics, Vol.QE 2, Sept.66, pp.235-242.

5. Hutchings, T.J. et al. — *Amplitude and Frequency Characteristics of a Ring Laser.* Physical Review, Vol.152, 1966, pp.467-473.

6. Dessus, B. et al. — *Effets Non Reciproques et Couplage dans un Laser en Anneau.* Comptes rendus, Académie des Sciences, Paris, Vol.262, 1966, pp.1691-1694.

7. Luziau, R. — *La Mise au Point des Gyromètres à Laser aux Etats-Unis.* Doc. Air Espace, no.104, mai 1967.

8. Podgorski, T.; Aronowitz, F. — *Langmuir Flow Effects in the Laser Gyro.* Présenté à la Conference on Laser Engineering and Applications, Washington, DC, 6-9 juin, 1967.

9. Catherin, J.M.; Dessus, B. — *Travelling Wave Laser Gyrocompass.* Présenté à la Conference on Laser Engineering and Applications, Washington, DC, 6-9 juin, 1967.

10. Killpatrick, J. — *The Laser Gyro.* Institute of Electrical and Electronic Engineers, Spectrum, October 1967, pp.44-54.

11. Heer, C.V. — Physical Review, Vol.134, 1964, p.A 799.

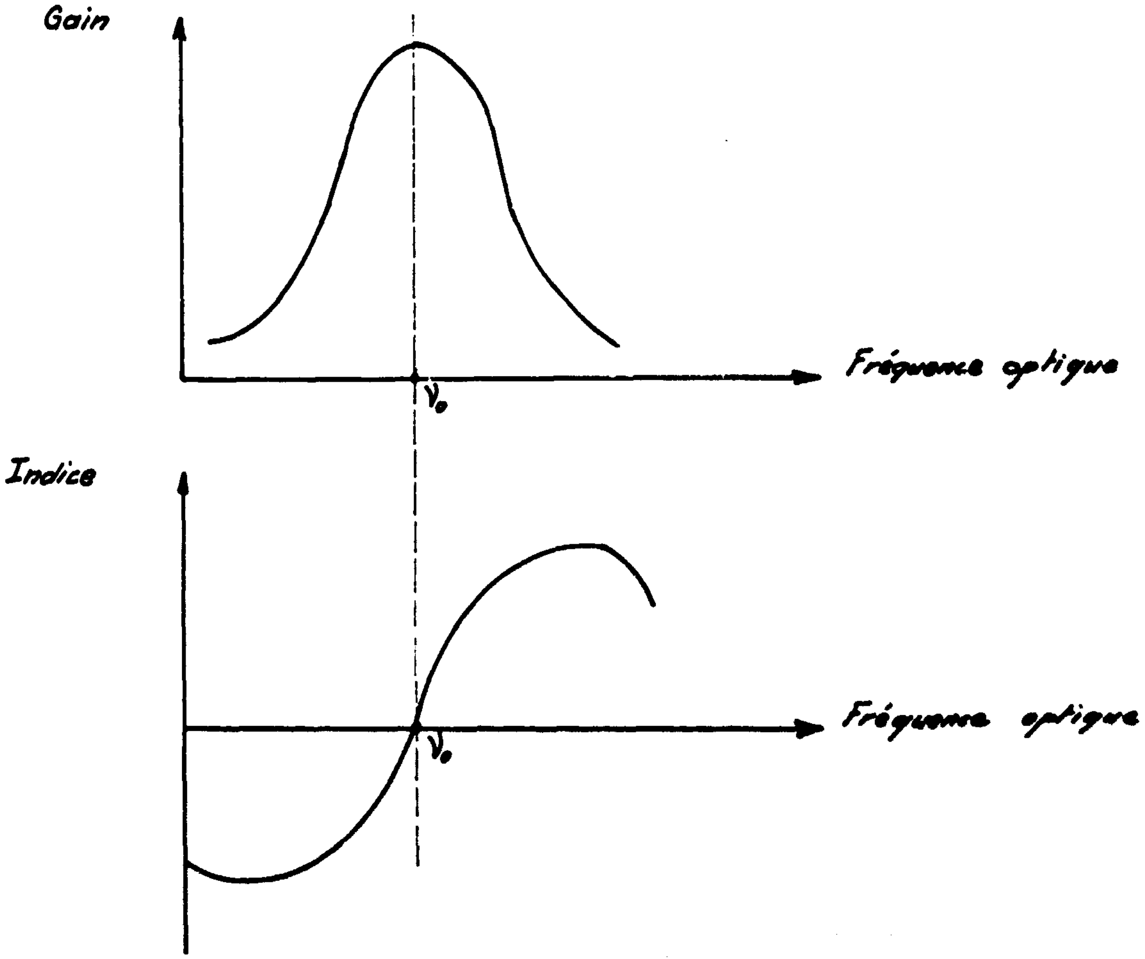


Fig.1 Variation du gain dans la raie d'émission et variation associée de l'indice (compte non tenu des effets de saturation)

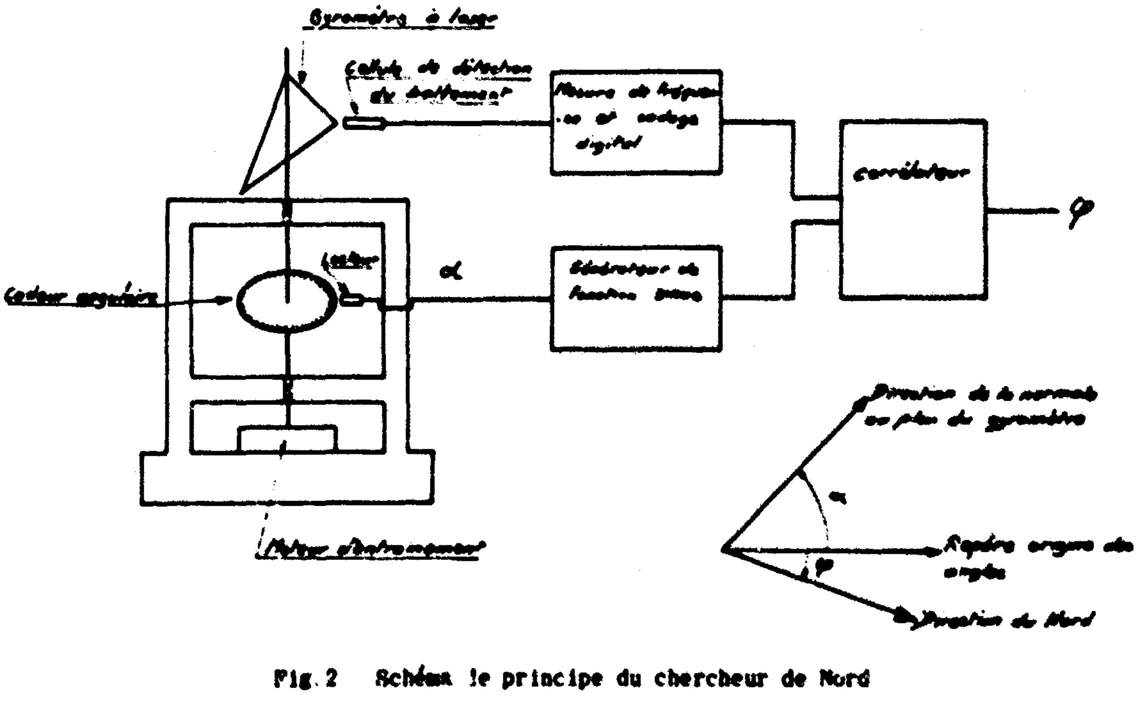

Gyromètre à laser
Cellule de détection du battement
Mesure de fréquence et codage digital
corrélateur
φ
Lecteur
α
Codeur angulaire
Générateur de fonction sinus
Moteur d'entraînement
Direction de la normale au plan du gyromètre
α
Repère origine des angles
φ
Direction du Nord

Fig.2 Schéma de principe du chercheur de Nord

STATUS AND FUTURE OF FLUID COMPONENTS

by

D.L.Wright and W.G.Wing

Sperry Gyroscope, Bracknell, Berkshire, England

STATUS AND FUTURE OF FLUID COMPONENTS

D.L.Wright and W.G.Wing

1. INTRODUCTION

1.1 General

The majority of inertial instruments produced today and over the last two decades have employed a rigid body, either spinning or otherwise, as their sensitive element. However, many of us have reflected on alternative approaches to this problem and to the problems involved in constructing such instruments. A mass of fluid is potentially attractive as a gyroscopic element since designs can be envisaged which require no mechanical bearings for the support of the spinning element. Furthermore, precessional torques affecting the direction of the momentum vector can be sensed by means of pressure transducers which are non-critical from a mechanical design point of view. Acceleration-sensing devices can also be envisaged using both spinning fluids and pressure or capacitive sensing techniques.

While the basic concept of such fluid instruments is simple, one does not obtain a usable instrument without the usual long drawn out battle with nature.

The increasing interest throughout the past decade in fluid sensors of a gyroscopic and acceleration sensitive nature arose initially as a result of possible applications in the stringent high g environment of short range, high acceleration missiles. Early designs and applications of such fluid components are described in this paper.

Particular examples of fluid gyros and accelerometers are described and their performance analysed.

The most commonly discussed fluid gyros are those employing the angular momentum of a spinning body of fluid. The fluid body in the most simple configuration is contained within a spherical cavity internal to a spinning case. The basic time constant of such gyros determines whether the instrument is used in either a "rate" or "displacement" mode. A further group of fluid gyros are those depending on the direct measurement of Coriolis acceleration. In these instruments the input angular velocity with a reference to inertial space is sensed by measuring Coriolis forces constraining a fluid mass relative to the rotating base of the instrument. It is possible to envisage instruments generating either translational momentum (oscillatory or otherwise) or rotational momentum to sense such forces.

Various types of fluid accelerometers have been designed, using the fluid itself either as the sensitive mass or as the sensitive mass-supporting medium.

The information presented here, on the range of sensors described, has been collected from various sources, including the Bendix Research Laboratories, Aviation Electric Limited, the National Water Lift Company*, the Douglas Aircraft Company Inc. and the Sperry Gyroscope Company, together with information available from various patents and publications.

---

* A division of the Pneumo Dynamics Corporation, part of Natronics.

1.2 Why Fluid Instruments?

The main reason in the past for considering the fluid instrument approach has really arisen from environmental considerations. Extreme conditions of shock, vibration and temperature have been typical of the early applications which were conceived. Extremely low g and g^2 sensitivities can be obtained in certain fluid gyroscopic designs. Advantages claimed for such devices are short reaction times, high over rating capability during gyro run-up, the possibility of eliminating temperature controls and reducing calibration requirements. Also, long-term storage capability, radiation resistance, bearings external to the sensitive elements eliminating certain mass shift effects, high reliability due to low bearing speeds and ability to effectively lubricate bearings all assist in obtaining a highly reliable device.

Accelerometers using fluid as the sensitive element have also been developed for reasons similar to the above, ruggedness in the face of stringent environments. The sensing system can be made non-pendulous and inherently isoelastic, resulting in very low g/g cross-coupling and g/g^2 errors.

1.3 Status and Applications

The development and performance of some of these fluid sensors has reached the stage where applications in the following fields may be considered, for some of the following reasons

Tactical Missile and Aircraft Guidance - High accuracy and inherent g-insensitive performance combined with low cost, simplicity, maintainability and ruggedness.

Stabilisation and Control - Wide dynamic range, low rate threshold and good dynamic response. Redundancy can be obtained by using two-degree-of-freedom rate devices.

Special Applications - High g designs with little or no trade-off in accuracy can be achieved, such as are required for anti-missile missile applications. Certain devices have been developed suitable for space vehicle stability applications.

Further developments into the performance range suitable for longer term applications are now envisaged.

2. TYPES OF FLUID INSTRUMENTS

Fluid Sphere Gyros - Using the angular momentum of a spinning sphere of fluid. The spherical cavity containing the fluid is sometimes completely filled, in some designs only partly filled. The fluids used have ranged from low viscosity fluorochemicals to mercury.

Fluid Tube Gyros - Employing a spinning closed hollow tube filled with liquid. Diaphragms across the hollow tube restrict the flow of liquid around the tube. A pressure is produced across the diaphragm as a result of Coriolis acceleration of liquid in the tube. The acceleration is similar to the Coriolis acceleration acting on the particles of a conventional gyro wheel. A similar device can be envisaged using a rotating annular body of mercury.

Vortex Rate Sensors - In such devices the fluid used has normally been in the gaseous state, to which a measure of the angular rate of the instrument housing is transmitted. The change in tangential velocity imparted to the gaseous stream causes the latter to spiral inwards towards axis of rotation. The change in tangential velocity is amplified as the gas moves inwards towards the rotation axis, due to the conservation of angular momentum. Various techniques have been proposed to sense the changes in radial and tangential flow as a measure of the applied angular rate. A synonymous liquid device has been proposed, the spiral stream of fluid driving a paddle wheel free to rotate about the input angular rate axis.

Vibration Gas Rate Sensors - A longitudinal sonic standing wave is maintained in a cylindrical cavity. Rotation about an axis transverse to this cavity is measured by pressure sensors such as microphones placed in the cavity on either side of the nodal point of velocity. Such instruments are another example of Coriolis acceleration sensors.

Free Surface Fluid Accelerometers - Sense the changes in shape of the free fluid surface internal to a spinning container partially filled with fluid. A three-axis device can be constructed. The fluid is the actual test mass.

Accelerometer with Fluid Suspension - Two- and three-axis units in which fluids are used to dynamically support a solid test mass with reduced or near zero suspension friction.

These and various other approaches to the problem of measuring angular rates and linear accelerations are described in more detail below. The possibility of combining a capability for measuring both types of input, accelerations and angular rates, within one instrument is mentioned.

3. GYROS

3.1 Fluid Rotor Instruments

Since the early 1960's several organisations, mainly in the USA have been working on such instruments. Advantages claimed for them, compared with solid rotor devices are as follows:

Accurate mass balance, since a uniform body of fluid has its point of support or centre of pressure at its centre of gravity.

Inherent isoelasticity if the fluid body is homogeneous

Absence of mass shift due to wear or bearing movement.

Absence of magnetic effects and torques induced by external magnetic fields.

Absence of rotor gimballing.

Absence of coulomb friction effects

Reduces signal and power transmission problem from sensitive element (i.e. absence of slip rings on drift-sensitive axes).

The requirement for the fluid to be homogeneous introduces a possible problem on such instruments, that of the appearance of thermal gradients across the body of the fluid resulting in a movement of its centre of gravity. If this effect occurs, mass balance errors are produced

Two possible methods of driving the fluid body are available. Rotation may be imparted to the fluid body by rotation of the case (mounted on spin bearings) containing the fluid, by means of an external motor. Alternatively, by employing a conductive fluid, rotation may be imparted by eddy current coupling with an external, rotating magnetic field. In the former method there is no dissipation of energy in the fluid, reducing possible thermal effects at the expense of additional moving parts, i.e. spin bearings. In the latter method of imparting rotation, dissipation of energy in the fluid body occurs due to viscous shear effects between the stationary case and moving fluid. This can result in thermal gradients in the fluid body, however, on the credit side the number of moving parts is reduced by the absence of spin bearings.

Fluid rotor gyros driven via case motion are in the class of instruments utilising the angular momentum of a spinning fluid mass - fluid gyros in which a rotation is imparted to the fluid by an external rotating magnetic or electrical field which reacts and couples with the fluid are usually in the class of instruments configured to measure Coriolis accelerations.

Combination of these two methods of imparting rotation can be envisaged but has usually been ruled out on account of increased complexity.

3.2 Fluid Sphere Gyroscope

3.2.1 General Principles

Such an instrument is illustrated in Figure 1 - consisting of a cavity completely filled with fluid of low viscosity (1 centipoise at room temperature). In fact, a fluorochemical was used. The spherical cavity is rotated on bearings by the external motor causing the liquid to rotate with it at the same speed.

In the absence of viscous shear forces the spinning fluid would generate an angular momentum vector which, once established, would maintain its spatial direction for small space rotations of the case. However, viscous shear forces are present, since relative motion occurs between fluid and case for any relative angular displacement between the spin axis of the fluid and the case. The shear forces produced result in a torque on the fluid body, precessing it into alignment with the spin axis of the case. Relative angular displacement is a transient effect, decaying exponentially according to the transfer function

$$K \frac{T_s}{1 + T_s} , \tag{1}$$

where

$$T \simeq \frac{R}{5}\sqrt{\left(\frac{2}{\nu\Omega}\right)} \tag{2}$$

and

T = time constant (sec)

R = radius of the spherical cavity (cm)

Ω = spin rate (rad/sec)

ν = kinematic viscosity (stokes).

The device may thus be regarded as either a rate gyro with a delay in its response or as a displacement gyro for high-frequency inputs and a rate gyro for long-term or steady-state inputs.

The liquid rotor acts much like a solid ball having a diameter slightly smaller than that of the cavity. This ball is separated from the cavity walls by a thin lubricating layer of the liquid. It is evident that to make use of the gyroscopic properties of this "free rotor" some means must be provided to sense the displacement angle between the spin axis of the cavity walls.

Figure 1 indicates ports through the cavity wall at points about 45^0 removed from the spin axis of the cavity. It will be seen from this figure that, if an angle exists between the two spin axes, the ports are at unlike radii with respect to the spin axis of the liquid body, and for this reason the centrifugally induced pressures at the two ports are unequal. One half-revolution after the indicated condition, the two ports will have switched positions so that the differential pressure will have reversed sign. An alternating pressure, therefore, exists between the ports. A diaphragm placed in the connecting passage and arranged as a deflecting element in a capacitance bridge will sense a pressure varying at spin frequency.

The output of the gyroscope is an a.c. voltage at spin frequency and has a phase angle which is a measure of the direction of the input rate in the plane normal to the spin axis. For this reason it is necessary, for proper signal interpretation, to demodulate the signal, using a reference voltage derived from an alternator coupled to the spin drive shaft. This alternator is built into the gyroscope.

By making use of two demodulators and two quadrature phases of the reference supplies, the signal can be broken down into two components which represent rates about a pair of orthogonal input axes. The simple electronics necessary are shown schematically at Figure 2. One gyroscope thus does the work of two single-degree-of-freedom gyroscopes.

3.2.2 *Design Parameters*

A large sphere is desirable, since both the time constant and the generated pressure increase as sphere size increases. In one such instrument, the SYG 2550 developed by the Sperry Gyroscope Company, a 1-inch diameter sphere was the largest that could be used without sacrificing other important gyro element designs such as the symmetrical placement of the waistband transducers.

Spin-speed selection is based on sensitivity, noise, and stability. The optimum speed for this gyro is relatively low as compared with solid rotor gyros. Although the generated pressure is proportional to the square of spin speed, the time constant decreases with the square root of speed. Also, vibrations are more pronounced at higher speed, and temperature gradients are greater. The present design operates at 6000 r.p.m.

As already stated, the generated pressure results from a displacement between the fluid spin axis and the cavity spin axis. For a step input of case displacement about an axis normal to the paper (see Figure 1), the generated pressure in dynes/cm^2 is given by

$$P = \rho\Omega^2 R^2 \sin\delta \cos\delta \sin\Omega t \ , \qquad (3)$$

where ρ is the fluid density in gm/cm^3 and δ is the angular displacement in radians. For small values of δ ,

$$P = \rho\Omega^2 R^2 \delta \sin\Omega t . \qquad (4)$$

This differential pressure, sensed by the transducer, goes to zero when the displacement angle goes to zero. Because the differential pressure is alternating it is detected easily, and a static shift of the position of the transducer diaphragm produces no output signal. For this reason there is no hysteresis error or time-dependent bias error due to the transducers. Also, the null of the alternating pressure coincides with the zero displacement between the spin axes, even if the sensing ports are at unequal angles with respect to the cavity spin axis.

3.2.3 *Environmental Capability*

The device has been subjected to 100 g shocks, sustained linear accelerations to 40 g, sinusoidal vibration up to 90 g peak and acoustical noise to 136 dB. The g^2 sensitivity was found to be 0.03°/hr g^2 and the basic transfer function of Equation (1) was verified under constant amplitude angular vibration conditions from 0.1 c/s to 200 c/s.

The diametrically opposed diaphragms around the waistband of the sphere were matched as closely as possible to the density of the liquid; hence they were essentially floated and almost insensitive to vibration. Furthermore, with the two transducer outputs being in phase, opposition under vibration additional cancellation is achieved, the diaphragms bow in the same direction for axial vibration and in opposite directions under the effect of the desired rate input.

3.2.4 *Anisoelasticity*

The theoretical anisoelastic coefficient of the fluid sphere gyro is zero. The liquid used is compressible and, as a result, its centre of gravity will shift under the influence of acceleration. However, for a perfect sphere, this shift is always directly along the total acceleration vector. For an ellipsoidal cavity, there will be an extremely small drift error given by

$$\dot{\theta} = \frac{-K\rho_0^2 a_x}{I_0}\int xy\ dV\ , \tag{5}$$

where

$\dot{\theta}$ = drift rate

K = compressibility coefficient

ρ_0 = nominal liquid density

a_x = acceleration

I_0 = moment of inertia of nominal sphere

x, y = principal axes.

The term $\rho_0 \int xy\ dV$ is the product of inertia for the body and is zero for all axes of a sphere. For an ellipsoid it is zero if x and y are the principal axes, but not for other axes.

For an ellipsoid that is almost spherical, the product of inertia about an axis at 45° to the principal axes is approximated by

$$|I_{xy}| = |\epsilon I_0|\ , \tag{6}$$

where ϵ is the eccentricity. It is thus found that the drift rate from this source for an ellipsoidal cavity is given by

$$\dot{\theta} \simeq K\rho_0 a_x^2 \epsilon/\Omega\ . \tag{7}$$

If Equation (7) is evaluated for the fluid sphere's operating constants, the drift rate is 3.7×10^{-2} deg/hr g^2, so that, even for a very crude cavity (i.e. $\epsilon \lesssim 0.01$), the drift from this source will be of negligible magnitude.

3.2.5 Linear Acceleration and Mass Unbalance

Mass unbalance of the fluid sphere gyro is inherently low but is affected by thermal gradients along the spin axis. Experience has shown that a gradient less than 0.01°F/cm is obtained under extreme environments. If this gradient appeared in a liquid ball, drift of 0.8°/hr g would result.

3.2.6 Performance

Figure 3 shows a typical performance trace for the Sperry SYG 2550 gyro showing two axis outputs in cardinal orientations and under sinusoidal exposure of input axes to earth rate. The trace begins at the bottom. Recorder zero reading is measured first. Next the gyro is oriented with spin axis vertical; outputs are recorded for axis 1 North and South, then for axis 2 East and West. The gyro bias is obtained by comparing the mid-point of these readings to recorder zero for corresponding axis. The gyro is next oriented with spin axis horizontal. Outputs are recorded for axis 1 vertical and axis 2 East. Acceleration sensitivity is obtained by comparing these readings with those obtained with the spin axis vertical. A calibration for the trace is shown near the top. The gyro spin axis is vertical and the rate table is driven at ½ deg/sec, sinusoidally exposing the gyro's input axis to the earth rate's horizontal component, 22.6 deg/hr peak-to-peak. The swing of gyro output shown for axis 2 is 13 divisions; one division is 1.74 deg/hr. Using this calibration, average bias and g sensitivity for the two axes are found to be 0.7 deg/hr and 1.0 deg/hr g . The rate and displacement threshold for the test gyro can be found from the trace. Threshold is the minimum output about null that provides the expected output using nominal scale factor. Within limits of the measuring techniques this is less than 0.05 sec of arc.

3.2.7 Reliability

From a reliability point of view, a fluid sphere has several advantages over a typical floated gyro unit. In a typical system, where three single-axis floated gyros are required, two fluid spheres will suffice, like any two-axis device. In the floated units, the maximum life of spin bearings is limited and failures are increased because of the higher rotational speed (12,000 to 24,000 r.p.m.), the fact that smaller bearings are required, theoretical lubrication requirements (necessary to avoid unbalance of the float), and the higher operating temperature of the bearings as compared to the fluid sphere.

The bearings on the fluid sphere gyro are external to the sensitive element and therefore position shifts in those bearings are not important. The bearings can be large and abundantly lubricated; furthermore, the optimum spin speed is low, 6000 r.p.m. or less, leading to lower failure rates and longer life.

In the fluid sphere gyro, the spin motor is external to the gyro case, and no delicate power leads (flex leads) are required. The failure rate of the spin-motor stators should be much lower than in floated gyros because weight and size are not critical; the motor is relatively large, it is easily cooled, and does not dissipate heat directly in the sensitive element. In addition, the stators use wire with heavy insulation and the end turns are not highly compressed.

Assembly and quality control limitations in floated gyros lead to such problems as fluid contamination, seal leaks and bubbles, cracked jewels, and critical machining operations. In the fluid sphere gyro, seal leaks and bubbles are of some concern, but critical machining operations and complete freedom from micro-contaminants are not of primary importance. The fluid body is in inherent mass balance, regardless of the actual shape of the cavity, and contaminants (heavier than the fluid) will be centrifuged to the outer periphery of the sphere, which is not the primary area of interest. As previously mentioned, the fluid sphere gyro has no float and, consequently, no jewels or pivots. This, from a reliability point of view, gives further emphasis to the parts count, which is favourably low.

The effects of the features on an actual reliability assessment are shown in Table II. The mechanical assembly of a fluid sphere gyro is compared to that of a representative floated gyro in terms of parts population and failure rates. For a platform with three sensing axes, the reliability figure of merit using two fluid sphere units is more than four times that obtainable with three conventional single-axis floated gyros.

The approach described has been pursued by the Sperry Gyroscope Company and actual hardware has been produced. The performance of such a device is shown in Table I indicating the typical characteristics that are designed for. Development work along these lines has also been carried out at Autonetics and Martin-Orlando.

3.3 Fully Filled Cavity with Float

An alternative way to sense the angular displacement between the spin axis of the fluid and the container is to place a low density rigid float in the fluid body, centred by the radial pressure gradients and forces in the spinning body. Axial restraint has to be provided by some other means, e.g. electro-magnetically. Electrical pick-offs can be used to sense angular displacement of a fluid body. Problems arising in this approach are that the axial position of the float must be precise, since the float represents an inhomogeneity in the fluid and, if axially displaced, can cause mass unbalance. If the pick-off senses displacement between the float and a non-rotating member and the container, spin axis can be a source of error.

A device of this kind was the subject of a patent issued to R.M.Smyth of Costa Rica in 1928, US Patent No.1890831. In America, the Arma Corporation have investigated this approach.

3.4 Partly Filled Fluid Rotors

Certain devices have been configured in which the angular displacement between spin axis and case is sensed by leaving a void in the centre of the fluid body. When spinning, the void in the fluid approaches a cylindrical shape and, in the absence of any input rate, the spin axis of the fluid and the container are coincident. Under the effect of an input rate the two spin axes diverge; the divergence can be sensed by a capacity pick-off. Problems arise with this type of instrument under axial acceleration conditions, the void becoming parabolic, leading to anisoelastic drift. Under acceleration normal to the spin axis the void becomes eccentric with respect to this axis and the pick-off could mistake eccentricity for angular displacement. Work was carried out along these lines in the late 1950's at the Sperry Gyroscope Company.

An alternative approach using a partly filled cavity is that taken by the Pneumo-Dynamics Corporation, part of Nortronics. A cavity is partly filled with mercury rotating at a high velocity. The mercury is contained in a spherical cavity supported by spin axis bearings and is driven by an external two-phase synchronous motor. Piezo-electric pick-offs are used in tendem, situated normal to the spin axis. The spin axis shaft passes through the sphere; small paddles jut out from the shaft into the body of mercury and there are piezo-electric (ceramic) elements at the base of the paddles. Coupling of the angular momentum vector with the input rate results in the pick-offs generating a sinusoidal error signal proportional to the applied input rate. When the spin axis of the rotor case and mercury body are different the paddles bend minutely, stressing the piezo-electric devices, thus creating the output signal which is brought out on slip rings. The phasing of the sinusoidal signal relative to the case is proportional to the direction of the input; Figure 4 shows this system in schematic form.

A two-phase reference generator mounted on the spin axis permits the error signal to be discriminated and resolved into rate information along two orthogonal axes. Advantages claimed for this device are that the error signal being self-generating, does not require any external excitation, resulting in a low sensing threshold in the absence of pick-off quadrature and noise at null. Threshold is limited by the spin axis bearings because of the sensitivity of the piezo-electric crystal pick-offs. Devices are being developed replacing the ball-bearing suspension versions, which have now reached the hardware stage, by a gas-bearing suspension system.

As rate sensors these devices have thresholds of 0.5°/hr with a maximum rate of 500°/sec in a package less than 0.7 in. in diameter and 1.9 in. in length. The company is considering applications in stability augmentation of fire control equipment, platforms, antennas and flight control systems. Possibilities are being examined in the strap-down inertial field. Operation under stringent environmental conditions is envisaged; shock, vibration and mass unbalance are not thought to be problems. It should be noted that g^2 drift effects have to be carefully considered in partly filled devices. Theoretically it is possible to adjust these to zero by means of an exact degree of partial fill of the cavity. Typical performance characteristics of such devices have been published, as shown in Table I.

A second approach using a partly filled cavity is that which the Universal Match Company were working on in the early 1960's. Here again, the fluid used was mercury. When this device was spun-up the mercury formed an annular ring, inside the cavity, about the spin axis. Permanent magnets located internal to the cavity provided a diametrical flux pattern across the ring of mercury, which acted as the primary of a transformer. The container or case of the device acts effectively as the core of the transformer system with a stator winding positioned about the spin axis as shown in the schematic of Figure 5.

Any relative angular displacement or motion of the mercury ring relative to the permanent magnets, and hence to the container, results in a voltage being induced in the mercury, the primary of the transformer system, and hence an output voltage is obtained across the secondary winding. The case of such a device could be made of some soft iron material. Under an angular rate input orthogonal to the spin axis, the field and voltage induced into the mercury ring is oscillatory.

3.5 Vortex Gyros

3.5.1 General

Vortex gyros are one of a family of angular rate sensors which are currently being considered for missile, space vehicle, aircraft and torpedo applications. Complete fluid control and flight stabilisation systems can be envisaged to operate in environments where temperatures of 1000°F are likely to be encountered. The schematic shown in Figure 6 shows a stream of fluid injected through the porous case of a vortex chamber, about the longitudinal axis of which is applied the angular rate input of the carrying vehicle. Due to the conservation of angular momentum, the tangential velocity imparted to the fluid flow at the coupling element (porous case) is amplified as it spirals inwards towards the axial exit hole along the input axis.

The angular velocity increases as the radius decreases. Various methods of sensing the output have been employed. A hot-wire anemometer system could be employed as a very sensitive sensor of this change in velocity. As an alternative, pitot heads could be mounted in the exit orifice, arranged to measure the tangential velocity of the exiting gas, two heads being employed in anti-phase to sense whether the input is being applied in one direction or another.

Aviation Electric Limited of Canada have developed a device of this type generating differential pressure signals proportional to the angular velocity input. Such a device has no moving parts and may be operated with air or other commonly available gases. Good dynamic characteristics are obtained and both the magnitude and direction of rotation are sensed by means of a push/pull type of pick-off. High reliability and potential low cost are claimed as advantages for these devices which may be coupled directly into a fluidic amplifier to obtain the desired output signal levels. Table I shows typical operating characteristics for such a fluid vortex rate sensor. Thresholds of 0.05°/sec with a dynamic range of 3000 can be achieved. A complete fluid state flight stabilisation system has been a subject of study by the Bendix Corporation and the Flight Dynamics Laboratory of the Research and Technology Division of the US Air Force. Figures 7 and 8 show a schematic of the flight system, together with the digital rate sensor employed. The system is briefly described below.

3.5.2 Fluid State Flight Stabilisation System

The need for a dynamic range of 500:1 in rate sensing, necessitating a marked improvement in a signal-to-noise ratio for pneumatic fluid rate sensors and amplifiers, prompted the selection of a digital sensing technique. The system chosen is shown in block diagram form in Figure 7.

The direct use of the rate sensor pulse rate information would require very fast computation rates and was therefore abandoned in favour of a time unbalance or pulse width modulation approach. The modulator keeps the information stored in the time domain while it is amplified and converts it finally at the servo input amplifier to an analogue signal with appropriate filtering, as shown in Figure 7.

Servo feedback of elevon position employs a pneumatic bridge pick-off technique compatible with the fluid state summing amplifier. The vehicle rate information is fed directly to the latter from the compensation network. The vehicle rate signal is not amplified by any analogue devices, in order to avoid degradation of the signal-to-noise ratio. The rate sensor is outlined below, further details of this system may be obtained by reference to the paper by L.B.Tapling and J.F.Hall*

---

* *Progress of the US Air Force, Research and Technology Division - FDL Program. Synthesis of a Pure Fluid Flight Control System.* - AIAA/ION Guidance and Control Specialist Conference, Minneapolis, Minnesota, 16-18 August 1965.

3.5.3 Digital Vortex Rate Sensor

The digital vortex rate sensor, shown schematically in Figure 8, has an output which is a pulse train, the period of which is proportional to the rate input.

Operation of the vortex rate sensor can be divided into rate coupling, signal amplification, and pulse rate conversion functions. The rate coupling mechanisms transmit a portion of the angular velocity of the transducer housing to the fluid stream by means of a porous coupling element. The change in tangential velocity imparted to the flow at the coupling element is then amplified as it spirals inward towards the chamber exit hole, due to the conservation of angular momentum. The exiting tangential velocity varies the angle of the conically-shaped free jet leaving the vortex chamber. The proportional signal at the chamber exit, i.e., the cone angle, is converted to a pulse rate modulated signal by means of a specially designed pick-off. The pick-off generates a pulse train by charging a volume to some upper threshold, and then discharging it to a lower threshold. The lower threshold level varies with cone angle, and thus rate input, and specifies the duration of the pulse period, or the pulse rate. A typical output train illustrating the variation in pulse period due to rate input is shown in Figure 9.

Typical sinusoidal performance characteristics of a digital vortex rate sensor operating on nitrogen gas are presented in Figure 10.

To convert the rate sensor pulse output into a pulse width modulated or time unbalance signal, the rate sensor is biased by a tangential control jet which sets up a given biased or centre operating pulse rate. If this bias remains fixed, the addition of inertial rate will either increase the pulse rate or decrease it, depending on the direction of inertial movement. The time unbalanced rate concept requires that the bias should fully reverse its direction between each output pulse. For this type of rate sensor, the time required to fully reverse the bias should be less than 10% of the average period of an output pulse. Consider, for example, a time unbalanced rate sensor to be operated at a centre frequency of 100 pulses per second. Each pulse period is 10 milliseconds. Therefore, the time to completely reverse the bias should be less than 1.0 milliseconds. The bias reverse time is related to the rate sensor chamber filling time, and thus the chamber volume. A model designed for rapid bias switching is shown in Figure 11. In this configuration rate coupling and rate signal amplification are sacrificed to minimise the time to reverse the bias.

3.5.4 Time Unbalance Converter

Given a pulsating rate sensor whose output frequency (and interpulse period) is a function of both pneumatic bias and inertial rate, a pulse network form is required which converts the rate sensor frequency output to a time-unbalance-modulated pulse train.

In the basic network (Fig.12(a)), each sensor pulse (output C) switches the output of a complementing flip-flop (outputs A and B). The flip-flop outputs are fed back to opposing control inputs on the sensor to establish the sensor bias direction. Since the flip-flop switches at each sensor pulse, the sensor bias direction also changes after each sensor pulse. Since the time between pulses is established by the vector sum of the bias and rate induced angular moments, rate alternately adds to the bias for one pulse period and subtracts for the next. Only the direction of the bias is alternated (the magnitude is constant) so that with zero rates the alternate pulse durations are equal. The output of the network is the time-unbalance-modulated pulse train which exists at the flip-flop output (outputs A and B) (Fig.12(b)).

4. GYROS - CORIOLIS ACCELERATION SENSORS AS GYROSCOPIC INSTRUMENTS

Several instruments have been designed and tested based upon the direct measurement of Coriolis acceleration forces. Most of the fluidic gyroscopic instruments with non-rotating cases are dependent upon such a measurement.

4.1 Rotating Field Instruments

One such instrument is shown in Figure 13. An annulus of conductive fluid, such as mercury, is spun by the application of a rotating magnetic field applied to a conventional polyphase motor stator. Any input angular rate with a component about an axis orthogonal to the axis of spin of the mercury annulus will cause a pressure difference to appear between a pair of ports located at either end of the annulus. This pressure difference arises directly as a result of Coriolis acceleration effects. If the radius of the location of the pressure sensing ports is R and the axial length of the annulus of mercury is L, the pressure generated by Coriolis acceleration is

$$P = 2\Omega R L \rho \omega , \tag{8}$$

where Ω = spin rate of annulus
ρ = fluid density
ω = input angular rate orthogonal to spin axis.

Positioning two sets of pressure sensing ports at right angles allows input rates about two axes to be sensed.

While such a gyro is attractive from its over-all simplicity and potential life, due to the absence of moving parts, power input requirements limit the value of Ω. The d.c. output pressure is also difficult to sense accurately at very low levels. Thermal problems may also be met in such a design if there is a large power input to the mercury or any other fluid used.

In the early 1960's the Speidel Corporation worked along similar lines to the example above. Their approach was to effectively sense disturbance in the flow pattern of the fluid by application of an angular rate input. Such disturbances cuased the eddy currents generated in the fluid to assume an asymmetrical pattern, shifting the magnetic field pattern, which could be detected via pick-up coils located on the stator. The output voltage obtained was at the supply, or spin excitation, frequency.

As stated in the Introduction, an oscillatory angular momentum vector can be used in such devices if the spin of the fluid rotor in the above example is made oscillatory rather than a continuous motion. The pressure produced by Coriolis forces would then alternate at the oscillating frequency of the fluid. While making pressure sensing easier, the resultant spin momentum vector, or oscillatory momentum vector, is found to be lower, since the maximum value of Ω which can be obtained practically is reduced. Additional errors can occur in such a device if the pressure transucers also sense some part of the pressure required to cause the oscillatory flow about the annulus. Additional compensation techniques would have to be developed to overcome this problem.

4.2 Fluid Tube Rate Gyroscope

Such a device senses the Coriolis acceleration forces set up in a spinning tube of fluid by the application of angular rate input orthogonal to the axis of spin. Such a device, as described below, has been produced for the stringent environment of missile applications and has been found to withstand vibration and shock levels up to several hundred g.

Figure 14 illustrates the fluid tube gyro principle. It shows a closed hollow tube filled with a liquid and a diaphragm across a section of the tube blocking the flow of liquid around the tube. This assembly is mounted on bearings and spun by an external motor.

For any input rate acting on this device there will be a pressure ΔP produced across the diaphragm. This is given by

$$\Delta P = 2A\rho\Omega \sin(\Omega t + \phi) \ . \quad (9)$$

where A = area enclosed by centre line of tube

ρ = density of liquid

Ω = spin velocity

ω = input rate

ϕ = phase angle of output which indicates direction in space of rate.

The pressure is the result of Coriolis acceleration on the liquid in the tube. This acceleration is comparable to the Coriolis acceleration which acts on the particles of a conventional gyroscope wheel. The pressure is sinusoidal and causes an oscillation of the diaphragm, which is sensed as the oscillation of a microphone diaphragm, and results in an a.c. voltage at spin frequency. This output is obtained for an input about any axis normal to the spin axis and the angle ϕ is a direct measure of the space direction of the input rate. Use of a reference generator on the spin axis and phase sensitive detection of the signal provides two output signals, one for each of two orthogonal components of the input rate. This device is, therefore, a two-degree-of-freedom rate gyroscope.

The cutaway diagram of Figure 15 illustrates the fluid tube gyro. Four tubes, 90° apart, are shown with pressure transducers at their centres. The fifth tube, shown through the centre of the device, is the expansion volume for the liquid when the assembly is partially filled. Transfer of the spinning liquid between the tubes is by circumferential flow in the manifold volumes at each end of the tubes. This construction, rather than the basic model shown in Figure 14, is used in order to obtain a two-phase signal, phase rotation information and the cancellation of vibration induced errors. For a very high acceleration environment, the gyro may be completely filled with fluid. Expansion bellows can be included.

Typical electronics for processing the fluid tube gyro signal are similar to those shown in Figure 2. The design described results in an instrument which has a very wide range and low threshold. The Sperry Gyroscope Company has investigated gyros of this type.

4.3 Fluid Toroid Gyro

Another form of fluid rotor gyro is shown schematically in Figure 16. This device was disclosed in US Patent 2953925 issued in 1960 to T.Yeadon. Fluid is pumped around a toroid by use of two electro-magnetic pumps. The application of an angular rate input about a diameter of the toroid will cause precession torque about a diameter orthogonal to the previous one, resulting in a twisting of the torsion bars on which the whole device is suspended. The angle of twist of the torsion bars is a measure of the input rate applied. It is believed that the Boeing Corporation investigated this device, using electro-magnetic pumps. The intention of the design is again to improve life and allow operation in high temperature environments. It will be noticed that this device is a variation on the conventional rate gyroscope configuration.

4.4 Vibratory Gas Gyro

One final device utilising the effect of Coriolis acceleration on an oscillatory fluid body is that disclosed by Granquist in US Patent 2999389 in 1961. This is shown schematically in Figure 17. A straight tube contains gas, the flow of which is caused to oscillate at the resonant frequency of the fluid in the tube. The tube acts in a similar manner to an organ pipe. The acoustic driver could be a loud speaker type of device located at one end. Other fluids could be used and this device could be regarded as being analogous in a translatory manner to the oscillatory momentum version of the mercury annulus device previously described.

Rotation about the axis transverse to the cavity is sensed by pressure transducers, such as microphones, placed in the cavity on either side of the velocity nodal point. Alternatively, the pressure difference at these diametrically opposed points can be sensed by a signal diaphragm, as shown in the attached schematic. The pressure difference is given by

$$P_1 - P_2 = \Delta P = 2V\rho D\omega , \tag{10}$$

where
D = diameter of cavity
ρ = fluid density
V = instantaneous velocity of oscillating fluid
ω = input rate orthogonal to longitudinal axis.

5. FLUID ACCELEROMETERS

So far as it has been possible to ascertain, very few fluid accelerometers have yet reached the production stage or have, in fact, even been configured in the form of laboratory hardware. The four examples discussed below fall into the two main categories described in the Introduction. Those instruments in which the fluid is used to support the test mass, and those instruments in which the fluid itself is the sensitive element.

In general the former devices, those in which the fluid is used as the sensitive mass supporting media, appear to be more complex in design and coustruction than the latter devices described above and their gyroscopic counterparts. Advantages which are claimed for most of the fluid accelerometer acceleration sensing devices are as follows:

Long life capability, due to absence of sensitive mechanical support elements.

Multiple axis sensing capability.

Low cross-coupling and inherent isoelasticity eliminating g/g^2 type of errors.

5.1 A Horizontal Integrating Accelerometer

This device is a single-axis sensing instrument which eliminates one of the advantages claimed above. Figure 18 represents the essential mechanical elements of such an accelerometer. If the cylindrical chamber is spinning in a horizontal attitude the float, being lighter than the surrounding fluid, is radially centred. Hence the fluid is centrifuged to the outer diameter and the float is brought into alignment with the spin axis. The float thus has no physical contact with the structural elements, being supported entirely by the viscous fluid, thereby reducing friction levels to a very low or near zero value. If acceleration is imparted along the spin axis, an inertial force equal to the product of the acceleration multiplied by the effective mass of the float will act on the float. Since the float is buoyant, having a lower density than the fluid, its effective mass is negative and hence the inertial force tends to drive it in the direction of the acceleration rather than causing it to lag with respect to the outer case, as a massive body would do.

As the float moves it will be retarded by the fluid resisting any displacement due to its viscosity. This resistance represents a velocity damping force appearing in the form of a pressure difference between the end faces of the float and viscous shear or drag effects along the cylindrical surface; it only exists when the float is in motion relative to the outer case. For laminar flow both the retardation effects mentioned are proportional to the absolute viscosity of the fluid and to the flat velocity with the chamber. A simple equation can be written to describe the longitudinal motion of the float in response to a steady-state axial acceleration by equating the inertial reaction force with the viscous resistance force.

$$ma = K\mu V_F \tag{11}$$

where
m = effective mass of float (i.e. buoyancy)
a = applied acceleration along spin axis
k = constant depending upon float and chamber dimensions
μ = absolute viscosity of fluid
V_F = velocity of float within chamber .

Rearranging terms:

$$V_F = \frac{ma}{k\mu} . \tag{12}$$

This states that the float velocity is directly proportional to the applied acceleration, provided buoyance and viscosity remain constant.

Notice now the result of taking the time integral of each side of this equation over some definite period, from zero to t :

$$\int_0^t V_F \, dt = \frac{m}{k\mu}\int_0^t a \, dt . \tag{13}$$

On the left side, the integral of the float velocity will be a definite displacement X from the postulated mid-point starting position, while on the right the integral of acceleration will be the velocity component of the carrying platform parallel to the spin axis, so that

$$\text{float displacement } X = \frac{m}{k\mu} V , \tag{14}$$

where V is the platform velocity component.

Thus the displacement of the float along the spin axis from its neutral position is a direct measure of the change of platform velocity in the same direction. The constant of proportionality, $m/k\mu$, can be varied over many orders of magnitude by selection of chamber and float dimensions, float buoyancy and fluid viscosity. It is thus possible to design an instrument having any desired relationship between float displacement and platform velocity, according to the expected velocity range of the vehicle.

From the above it will be seen that the float position relative to its null with respect to the outer case represents platform velocity. Hence the float will only be in motion during a period of acceleration, a component of which is along the spin axis of the instrument. It will otherwise stay fixed in its displaced position as long as the vehicle velocity remains constant. The float will not return to its original starting point until the vehicle ceases to have motion parallel to the spin axis of the instrument. The instrument, in its final hardware form, utilises a variable capacitance pick-off as shown in Figure 19. The structural design is modified so that the float is annular in cross-section, centred about a central pillar or rod which carries one of the pick-off elements. The float is made of conductive material (aluminium) and the fluid is a dielectric, thereby allowing capacitances to be formed between the cylindrical surfaces of the float and the fixed case. As the float moves axially, two of the capacitances, C_R and C_L , vary linearly but oppositely with float displacement; the other two, C_0 and C_{FB} , are invariant with float displacement since they are always fully covered. The pick-off system is excited with a 400 cycle a.c. voltage. The float when displaced assumes a voltage, the magnitude of which is porportional to the displacement and the phase of which is dependent upon the direction of displacement from the central position. Figure 19 shows the introduction of a feedback loop to reduce coercive effects on the float arising by combination of the float voltage and the excitation voltage. The magnitude of the output voltage is given by

$$E_0 = E_i \frac{(C_R - C_L)}{C_{FB}} \quad . \tag{15}$$

Temperature effects have to be considered in this kind of device. A silicone oil was used for the flotation fluid because of its flat curve of viscosity versus temperature. The buoyancy term in Equation (13) also varies, since the relative densities of the float and fluid change with temperature. However, both terms in the ratio m/μ vary in the same sense with temperature changes. It is thus possible to specify a float density which will result in a rate of change of buoyancy equal to the rate of change of viscosity at some desired operating temperature. Very close control of float density is still required, however, to achieve compensation near the desired operating temperature and a heating element is also fitted to the instrument.

Variations in gaps between the float and the chamber, and between the float and central spool, can cause problems. Very close tolerances on run-out of the float have to be maintained, posing special problems in dynamic balancing. Tiny distortions in gap thickness caused by radial float displacement, or by a tapered float or bore, can result in an intolerable scale factor variation for the instrument. These sources of error are removed by means of by-pass channels for the fluid, from one end of the chamber to another. This technique also helps to overcome any creep of the float due to thermal gradients. The general principle can be understood very easily by referring to Figure 20 and imagining two extreme situations: firstly, only a central by-pass tube present (outer tubes blocked) and, secondly, only the outer channels present (the centre tube being a solid rod).

In the first case all the fluid moving to the left through the central bore would have to return through the float chamber, carrying the float with it to the right. In the second case the opposite would be true, the flow to the right through the outer tubes would return to the left through the float chamber, moving the float in that direction. Thus, for the same conditions of thermal gradient, opposite float motions could be induced. Evidently, by proper distribution of the flow between the inner and outer pathways, an equilibrium condition of zero net force on the float can be reached. The circulating fluid flow still exists, but all the axial forces on the float cancel one another, so that no erroneous motion results.

Such a device was constructed in the late 1950's by the Sperry Gyroscope Company. The performance of this device is shown in the typical parameters given in Table III.

5.2 A Three-Axis Integrating Accelerometer

Sperry have also configured a three-axis integrating accelerometer in which a single test mass is utilised to sense acceleration components along three orthogonal axes. The test mass is a 3 inch diameter sphere having positive buoyancy in a fluid-filled chamber within a $1\frac{7}{8}$ inch aluminium cube. Closely surrounding the sphere, but not touching it in normal operation, are three pairs of spherically contoured support pads, each pair being at opposite ends of a spherical diameter, the three pairs forming an orthogonal set. The instrument is shown in Figure 21.

An acceleration component along any of the three principal cube axes will result in an inertial displacement of the ball towards one or other of the two support pad faces associated with the axis considered. Ball displacement is sensed by a capacitive pick-off which detects any departure of the sphere from its mid-position, and transmits a signal of appropriate phase and magnitude to an external servo loop which activates a screw-type hydraulic pump. The pump, also housed in the cube (one for each axis), generates a fluid pressure differential between the pad faces, through a central perforation proportional to the rotational speed of the pump shaft, by virtue of the Newtonian properties of the silicone fluid employed. Since the pressure is also just sufficient to maintain the null position of the sphere against the acceleration-derived displacement force, pump r.p.m. is exactly proportional to the acceleration component along the axis. Therefore, shaft angular displacement represents the time integral of acceleration, or the total velocity along the axis.

Hydraulic interaction among the three axes is avoided by making the actual volumetric flow through the support pads extremely minute, most of the pump's output being circulated through isolated by-pass channels external to the ball chamber. These by-pass channels also serve the purpose of making the total hydraulic impedance of each circuit almost independent of possible variations in ball-to-pad clearance, such as might be caused by a slightly imperfect sphere. This sharply reduces the sensitivity of scale factor to such gap variations, and the small error which still could exist from this cause is almost totally eliminated by special compensating grooves in the pad faces.

Another important error compensation is achieved by selection of float and fluid characteristics so that, in the vicinity of the nominal operating temperature of the unit, the percentage change of fluid viscosity is exactly equal to the percentage change in float buoyancy for a small temperature change. Since scale factor (pump r.p.m. per unit acceleration) is proportional to the quotient of buoyancy divided by viscosity, it is thus made nearly invariant over a small temperature range. The actual extent of the variation is 0.0045% per $({}^{o}F)^2$, so that a control temperature of $\frac{1}{3}{}^{o}F$, which is reasonably attainable, results in only 0.0005% scale variation.

Positive coupling, between the pump servo motor external to the cube and the pump shaft itself, is accomplished by means of a pair of concentric cylindrical magnets, the inner one being contained within the hydraulic fluid and attached to the pump rotor. In this way no shaft extension pierces the wall of the accelerometer and no rotating seals are required. Thus the key elements of the unit (test mass, support pads, fluid and restoring pumps) are protected by the integrity of a perfect hermetic enclosure, while the drive motor, read-out device, and all servo electronics are externally mounted and accessible for maintenance. Figure 22 shows the system in schematic form.

Read-out is provided by a digitiser geared to each pump shaft, employing no slip rings. The digitiser is packaged in a Size 8 synchro case and produces 800 voltage pulses per pump revolution. A logic circuit switches this pulse train to a plus or minus output wire, depending upon the direction of rotation. The device was scaled for the particular application such that each pulse represented a velocity change of 0.1 ft/sec.

The principal design features and advantages claimed for this device are given in Table IV. It will be noticed once again that the principal advantages claimed for this approach are those of operational capability in rugged environments, leading to increased reliability.

The performance of this device is shown in Table II.

5.3 Free-Surface Accelerometer

Such a device employs the fluid directly as the sensitive element of the instrument. Figures 23 and 24 show the operation and layout of such a device in schematic form.

The free-surface accelerometer, in its most elementary form, consists of a spinning container partially filled with a dielectric fluid. In a zero-g field (Fig.23(a)), the centrifugal forces on the fluid maintain a free surface which is cylindrical in shape, with an axis which coincides with the spin axis of the container. The free surface of the fluid, however, forms a section of a paraboloid (Fig.23(b)) when the container is accelerated along its spin axis. The free surface of the fluid is also translated in the direction of the acceleration vector when this vector is normal to the spin axis (Fig.23(c)).

With acceleration in the direction of the container spin axis, the intercept of the paraboloid and any plane normal to the spin axis is a circle. Two such circles are formed by the free surface and the end plates of the container, and the difference in these two circular areas is directly proportional to the applied acceleration. The difference in the areas is given by

$$\Delta A = \frac{2\pi L a_z}{\omega^2} \tag{16}$$

where ΔA = difference in end circle areas

a_z = acceleration in z (spin axis) direction

L = length of fluid body

ω = spin rate.

When the components of acceleration in the plane normal to the spin axis of the container are a_x and a_y (x, y and z forming an inertial coordinate system), the components of displacement of the axis of the free surface relative to the spin axis of the container are, respectively ϵ_x and ϵ_y. These displacements, or eccentricities, are directly proportional to the applied acceleration and are given by

$$\epsilon_x = \frac{a_y}{\omega^2} \tag{17}$$

$$\epsilon_y = \frac{a_x}{\omega^2}. \tag{18}$$

In the presence of applied acceleration simultaneously occurring along and normal to the spin axis, the resultant fluid motions are superimposed; there is no cross-coupling.

Pick-off for Acceleration Along Spin Axis

By using a dielectric fluid, a capacitive bridge pick-off is devised to measure the area change of the end plate circles. This provides the output proportional to the acceleration component acting along the spin axis.

Pick-off for Acceleration Normal to Spin Axis

To detect the translation of the free surface in response to acceleration normal to the spin axis, a capacitive pick-off, consisting of two split discs and one spool, is employed. The discs are split into two halves along a diameter, and are placed at each end of the container. The spool, or element, is placed at the centre of the container. Each half of the split disc is excited from the secondary of a centre-tapped transformer.

Depending on the instantaneous value of the eccentricity, the amount of fluid encompassed between each split plate and the output disc is unequal. Thus the output disc is caused to assume a voltage whose magnitude and phase is proportional to the magnitude and direction of the eccentricity.

For a constant input acceleration, normal to the spin axis, the free surface will be displaced from the spin centre, and will remain constantly displaced in the direction of the steadily applied acceleration. Therefore, with respect to the rotating split-plates, the encompassed fluid seen by the individual plates will vary harmonically once per revolution of the drive motor. The output of the pick-off is thus an a.c. voltage at the spin frequency.

Use of a suitable reference generator on the drive shaft, in combination with the phase-sensitive detection, provides two output signals, each proportional to a cross-axis acceleration.

Advantages claimed for the device just described are absence of pendulosity, i.e. the sensitive mass is non-pendulous, and complete isoelasticity. The optimum spin speed of the container is found to be quite low, which leads to higher reliability and long bearing life; the latter bearings are also external to the sensitive elements and hence can be quite large and well lubricated, again leading to long life and allowing their selection for high acceleration environment. In the diagram of Figure 24 it will be noticed that inductive slip rings (air-gap rotary transformers) are employed, eliminating metal-to-metal contact slip ring and brushes, thereby again extending reliability. Such an instrument could have been built in the same container as the fluid sphere gyro described earlier, and the container spin rate imparted by the same motor to give a compact gyro accelerometer package; the gyro shaft reference generator could also have been used as a common element in discriminating the cross-axis acceleration components.

The main performance and design characteristics of the free surface accelerometer devices are shown in Table III. Figure 25 shows the signal-processing arrangement required.

From the practical point of view a low viscosity dielectric fluid is used and large gap spacings, outside the pick-off region, in order to reduce phase offset in the direction of spin due to viscous shear. The latter is caused by relative velocity between the fluid void boundary and walls of the container. This latter problem can be determined and, being stable, can be compensated for over a wide temperature range.

The hydraulic scale factor of the instrument is conceptually independent of temperature. However, in combination with the capacitive pick-off, the three axis voltage outputs are subject to certain temperature-dependent error sources. Specifically, these errors are caused by axial dimension changes in the pick-off, coupled with change of fluid dielectric constant as a function of temperature.

For the geometry of the three-axis prototype, the dimensional and dielectric constant changes each lead to scale factor changes of approximately 0.08% per ^{o}F. Fortunately, the internal feedback arrangement of Figure 26, in combination with certain passive networks, makes it possible to trade-off one scale value change versus the other, leading to temperature error reduced by a factor of approximately 10 to 1.

Thus residual scale value versus temperature errors will be less than 0.008% per ^{o}F and, for an anticipated control excursion of ± 20^{o}F, will contribute peak scale factor errors of less than ± 0.16%.

Such a device has been built and tested by the Sperry Gyroscope Company.

6. COMBINED ANGULAR RATE AND ACCELERATION SENSING INSTRUMENTS

Combined instruments can be envisaged; however, few have reached the hardware stage.

One approach which has been considered is the combination of a fluid tube gyro along the lines already described, with a free-surface fluid accelerometer. Examination of Figure 15 shows that the central chamber of the fluid tube gyro, if partly filled, would act as a free-surface accelerometer. Such a device has been considered by Sperry. Problems involved are that of signal sensing and extraction of the data from the sensitive element, without paying the penalty of excessive design and manufacturing complications.

7. CONCLUSIONS AND ACKNOWLEDGEMENTS

The author would like to thank all the firms and their personnel who have contributed information contained within this paper. Presentation of the latest state of the art has been somewhat restricted by Governmental and Company security restrictions. However, it is believed that enough has been presented to indicate that such instructions are now practical hardware propositions and could have wider applications than those in the purely military missile field for which they were initially configured. The main advantages which are likely to be claimed in the future, as the performance of these instruments approaches that of current solid rotor and conventional acceleration sensors, are those of low cost, high reliability, long life, together with freedom from g-sensitive error effects.

If mention of other organisations who are active in this field has been omitted, the authors wish to apologise and would welcome their additional contribution.

TABLE I

Typical Fluid Gyro Performance Details and Characteristics

Gyro	*Fluid Sphere Gyro (fully filled)*	*Fluid Sphere Gyro (partly filled)*	*Fluid Sphere Gyro (part filled – in development)*	*Fluid Vortex Rate Sensor*	*Fluid Tube Rate Gyro*	*Vibrating Gas Gyros (straight tube, laboratory tests)*
Size (in.)	$2\frac{3}{8}$ dia × $4\frac{3}{8}$ L	0.7 dia × 1.9 L	0.48 dia × 1.8 L	$6\frac{1}{8}$ dia × $1\frac{7}{8}$ L	$2\frac{3}{8}$ dia × $4\frac{3}{8}$ L	-
Weight	2 lb	80 gm	24 gm	2 lb	2 lb	-
Dynamic range	10^6:1	3.6 × 10^6:1	1.4 × 10^6:1	3 × 10^3:1 (linear)	10^5:1	-
Maximum rate	20,000°/hr	500°/sec	700°/sec	150°/sec (linear (300°/sec max)	500°/sec	-
Threshold	0.02°/hr	0.5°/hr	0.005°/sec	0.05°/sec	0.005°/sec	0.001°/hr (claimed)
Non-g-sensitive drift	0.2°/hr	-	-	-	-	-
g-sensitive drift	0.1°/hr g	10°/hr g	-	-	-	-
g^2-sensitive drift	0.01°/hr g^2	-	-	-	-	-
Random drift	0.02°/hr	-	-	-	-	-
Day-to-day repeatability	0.04°/hr	10°/hr	-	-	-	-
	0.02°/hr g	-	-	-	-	-
Time constant or natural frequency	0.2 sec	90 c/s	-	10 msec	-	-
Linearity (%)	-	0.5	0.5	1	0.25 – 0.5	-
Shock	50 g/11 msec	2000 g	-	-	50 g/11 msec	-
Vibration	20 g/15 to 2000 c/s 0.1 g^2 (c/s)/15 to 2000 c/s	- -	- -	- -	20 g/15 to 2000 c/s 0.1 g^2 (c/s)/15 to 2000 c/s	- -
Reaction time (sec)	10	3	8	-	10	-
Manufacturer	Sperry	National Water Lift Company	National Water Lift Company	Aviation Electric Ltd	Sperry	Patentee: C.E.Granquist

TABLE II

Reliability Comparison

Component	*Quantity (n)*		*Failures/10^6hr (f)*		*nf*	
	SDF Gyro	*Fluid Sphere Gyro*	*SDF Gyro*	*Fluid Sphere Gyro*	*SDF Gyro*	*Fluid Sphere Gyro*
Spin bearings	2	4	30.0a	7.0a	60.0a	28.0a
Windings:						
spin motor	1	1	15.0a	6.0a	15.0a	6.0a
heater	1	0	3.0a	–	3.0a	–
torquer	1	0	10.0a	–	10.0a	–
signal generator	1	0	8.0a	–	8.0a	–
excitation	1	0	4.0a	–	4.0a	–
magnetic coupler	0	2	–	1.0a	–	2.0a
sensor wire (heater)	1	0	1.0a	–	1.0a	–
Reference generator	0	1	–	2.5a	–	2.5a
Transducer	0	4	–	1.0a	–	4.0a
Bellows	2	2	3.3a	3.3a	6.6a	6.6a
Ceramic capacitor	0	4	–	0.01a	–	0.04a
Resistor	0	4	–	0.01a	–	0.04a
Connections	40	40	0.01a	0.01a	0.4a	0.4a
				Total	108.0a	49.58a
				Assembly and quality control factor (35%, 10%)	37.8a	4.96a
				Failure rate, 1 gyro	145.8a	54.54a
				Failure rate, 3-gyro system, 2-gyro system	437.4a	109.2a

Note: **Failure rates normalised about failure-rate parameter a**

TABLE III

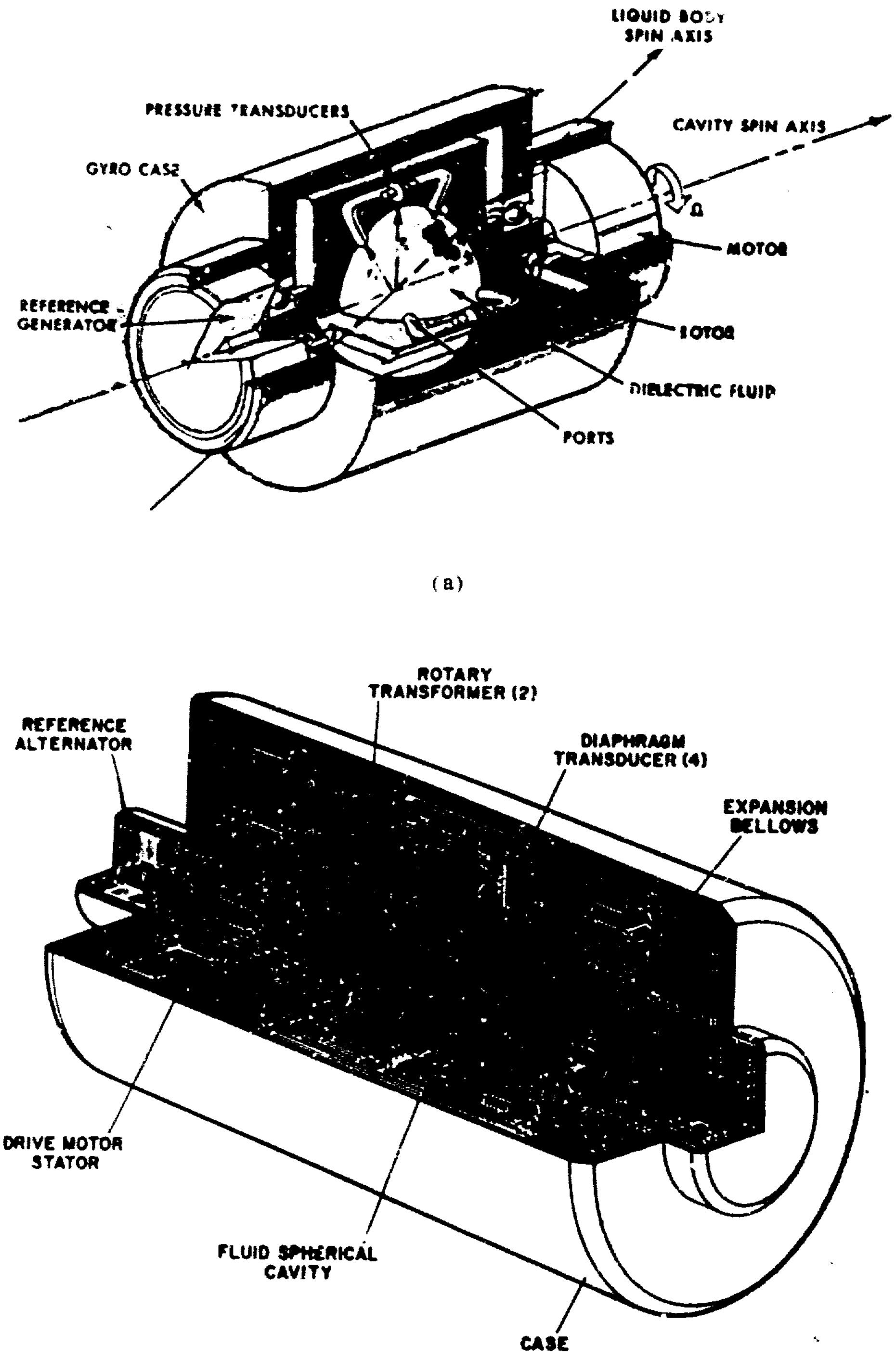
Typical Fluid Accelerometer Performance Details and Characteristics

	Three-Axis Integrating Accelerometer	*Free Surface Accelerometer*	*Magnetic Fluid Suspension Accelerometer (Design Aim)*
Size (in)	4 × 14 × 4 cube	$2\frac{1}{8}$ dia × $4\frac{1}{8}$ L	1 dia × 2 L
Weight (lb)	1.9	1.25	
Range	± 30 g	± 28 g	
Threshold	10^{-6} g	2×10^{-4} g	5×10^{-5} g
Linearity	3×10^{-4} g to 1 g 0.03% above 1 g	3×10^{-5} g to 1 g 0.003% above 1 g	
Scale factor	0.5 V d.c./g	100 r.p.m./g 0.1 (ft/sec)/bit	
Reaction time	-	10 sec	
Manufacturers	Sperry	Sperry	AVCO Corporation

TABLE IV

Principal Design Features of the Three-Axis Integrating Accelerometer

Feature	*Advantages*
Three axes embodied in a single compact assembly	Precision of orthogonality attained by basic design and manufacture; not a function of installation variables Small size and weight Lower cost than three single-axis units Extreme rigidity of small cube guarantees no deformations or relative axis misalignment under high g stresses
Null-type pick-off	Pick-off linearity not significant
Integration included in pump restoration function	No separate integrator package required
Linearity derived from viscous shear integration	Newtonian characteristics of flotation fluid result in inherently linear performance
Test mass heavily damped	Excellent vibration and shock resistance
Non-pendulous, non-mechanical test mass support	Extremely low threshold sensitivity No sensitive axis bias No cross-coupling (g^2) error No mechanical supporting elements to introduce spurious forces, coercion, etc No fragile suspensions subject to shock or vibration damage
Output in form of mechanical shaft rotation	Flexibility in choice of read-out elements (analogue digital)
Lightly loaded, slow speed bearings and gears	Negligible wear in thousands of hours of operation; extremely long life capability
No metal-to-metal rubbing contact in screw pump assembly	No performance deterioration due to wear; minimum self-heating of pump
Large percentage of aluminium cube structure remains after machining	Excellent heat distribution within instrument, minimising thermal gradients High structural rigidity; no significant g deflections
Few parts, rugged construction	Comparatively simple to manufacture and assemble



(a)

(b)

Best Available Copy

Fig. 1 Fluid sphere gyro. Non-floated rotor houses, liquid-filled spherical cavity. Other parts are outside fluid area

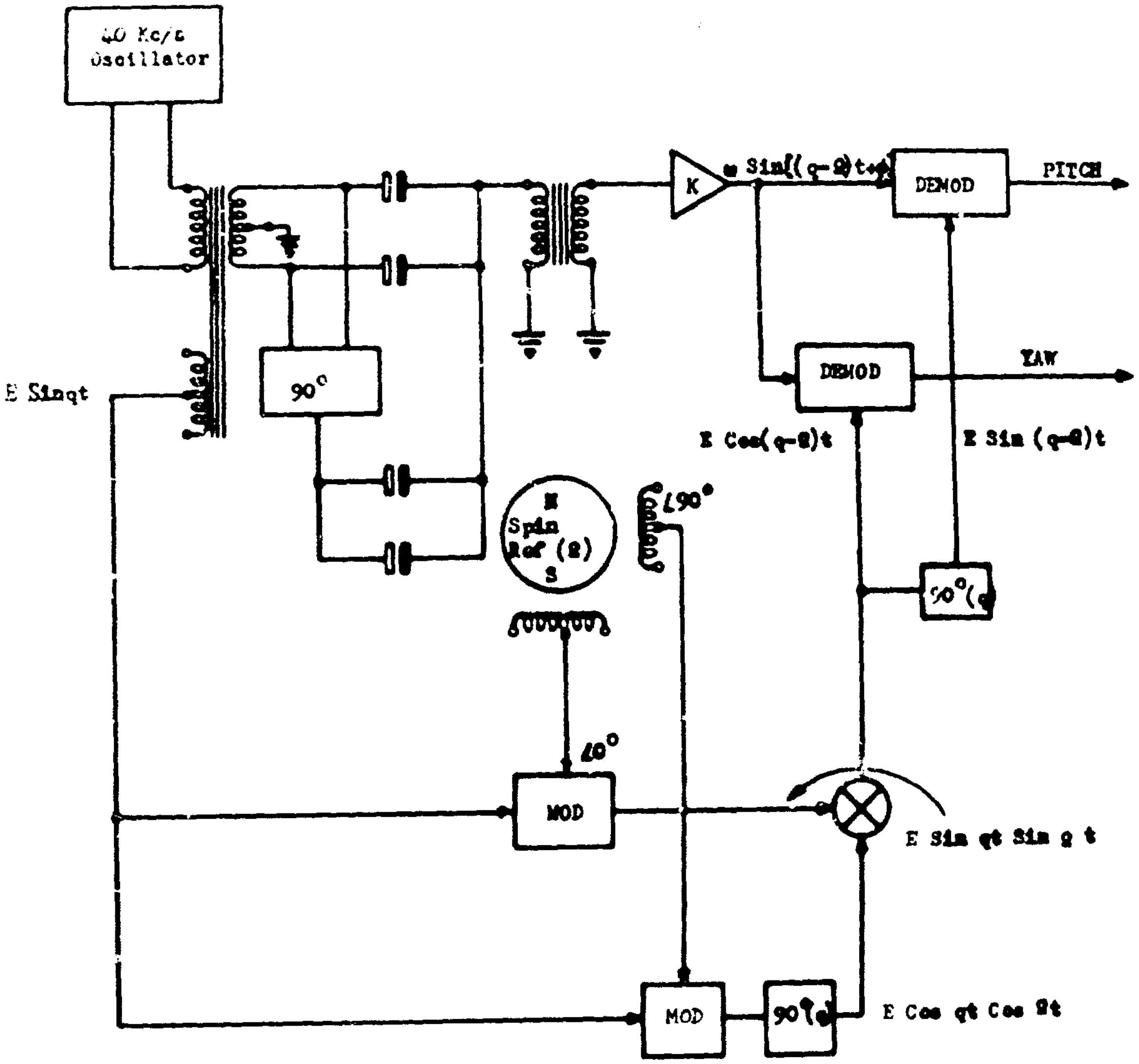


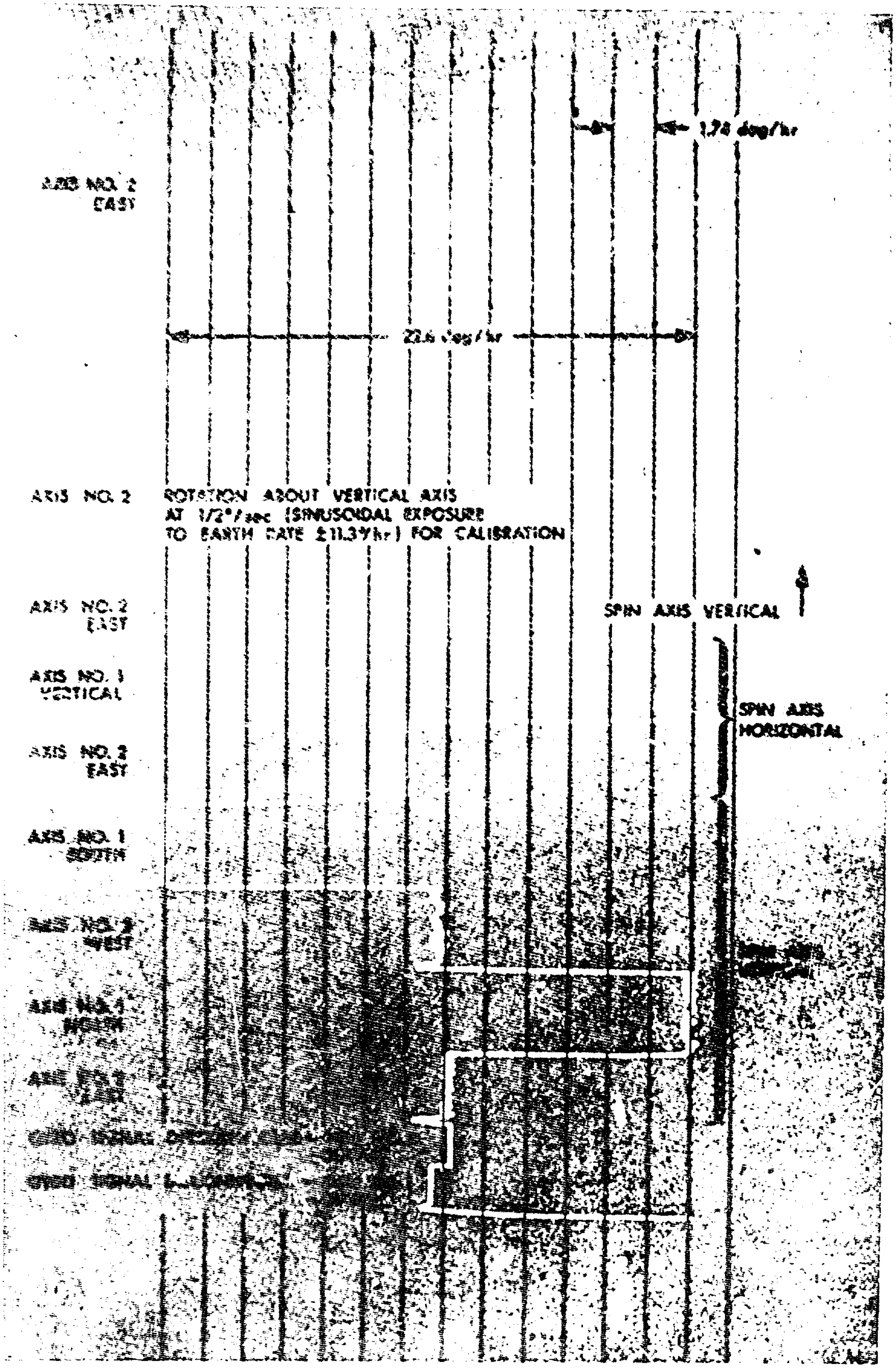
Fig. 2 Fluid sphere. Signal processing

ω = input angular rate ($= d\theta_1/dt = \delta/\tau$)

Ω = spin angular rate

q = carrier angular frequency

Best Available Copy



Best Available Copy

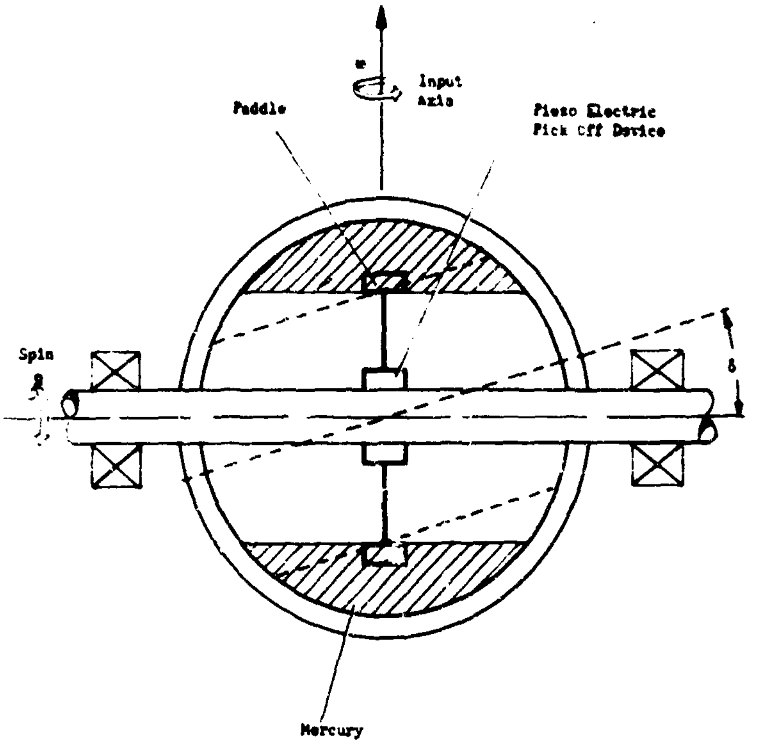
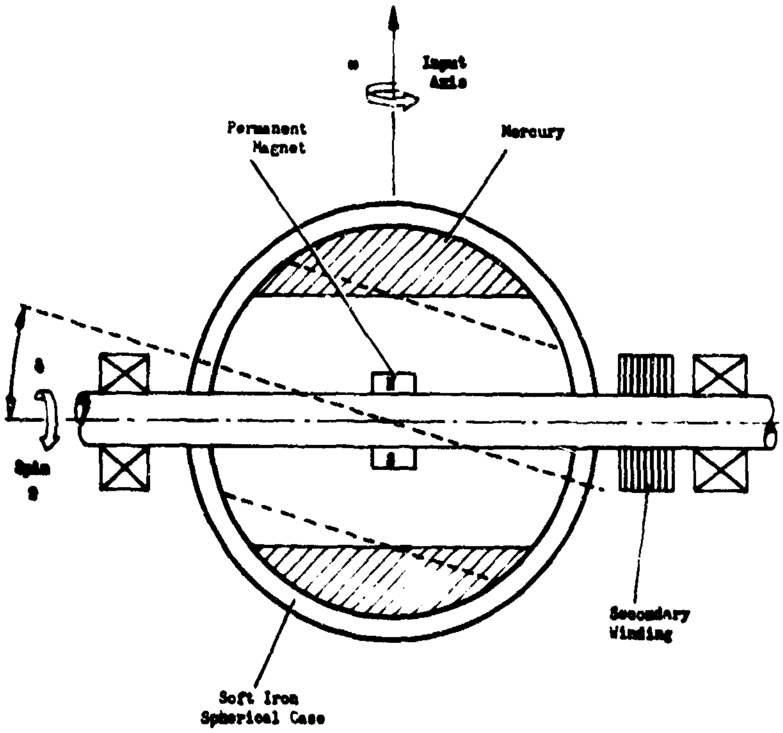


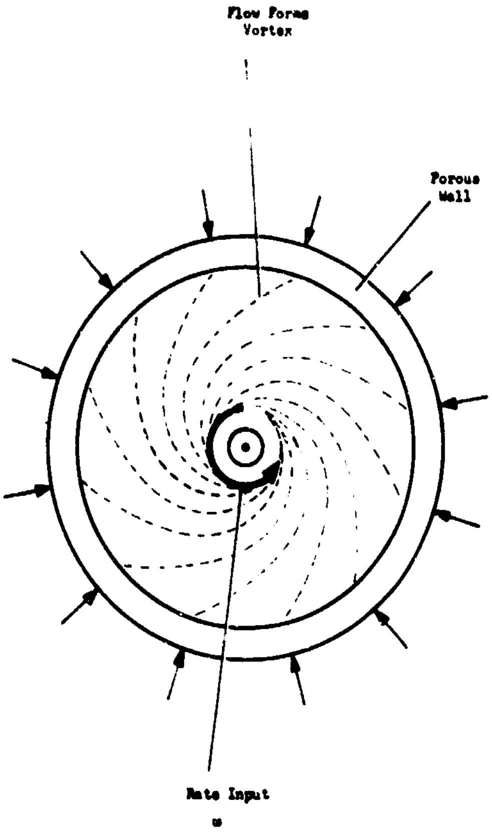
Fig. 4 Spin imparted by a two-phase synchronous motor



NOTE. 1. Permanent Magnet Provides Diametral Flux Across Mercury Ring.

NOTE. 2. Oscillatory Field If Relative Displacement δ Occurs Under Rate Input.

Figure 5



NOTE: Sensing Device Located Normal To Central Or Axial Exit.

Fig.6 Vortex rate sensor schematic

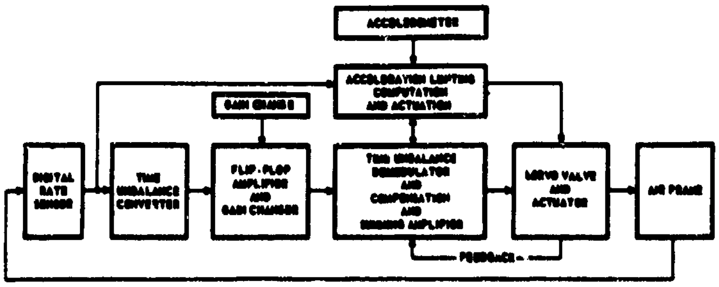

ACCELEROMETER
ACCELERATION LIMITING COMPUTATION AND ACTUATION
GAIN CHANGE
DIGITAL RATE SENSOR
TIME UNBALANCE CONVERTER
FLIP-FLOP AMPLIFIER AND GAIN CHANGER
TIME UNBALANCE DEMODULATOR AND COMPENSATION AND SUMMING AMPLIFIER
SERVO VALVE AND ACTUATOR
AIR FRAME
FEEDBACK

Fig.7 Block diagram of fluid state flight stabilisation system

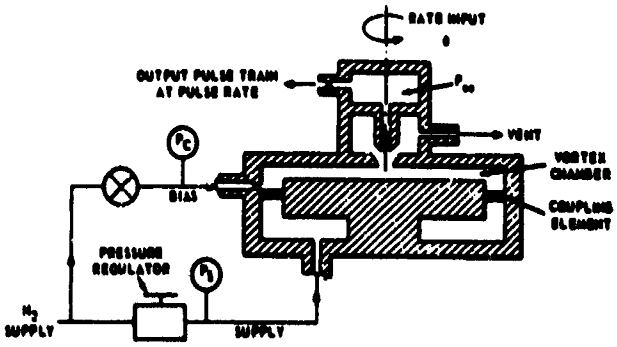


Fig.8 Digital rate sensor schematic

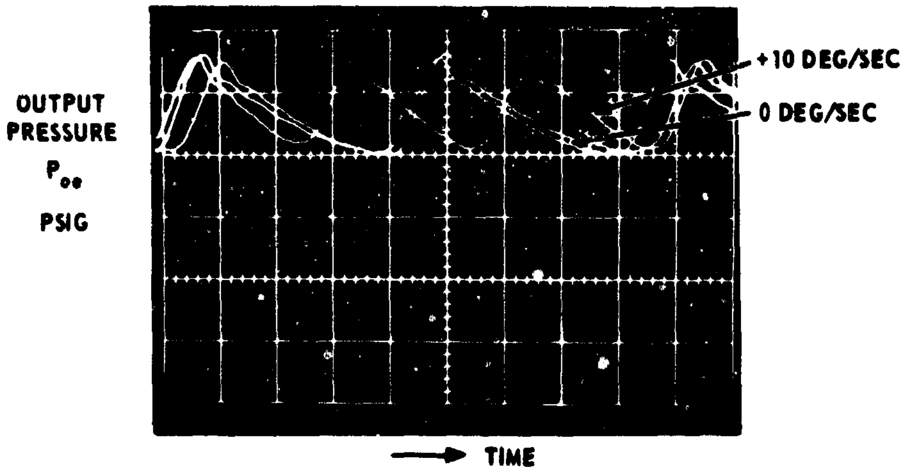


Fig. 9 Output pulse train

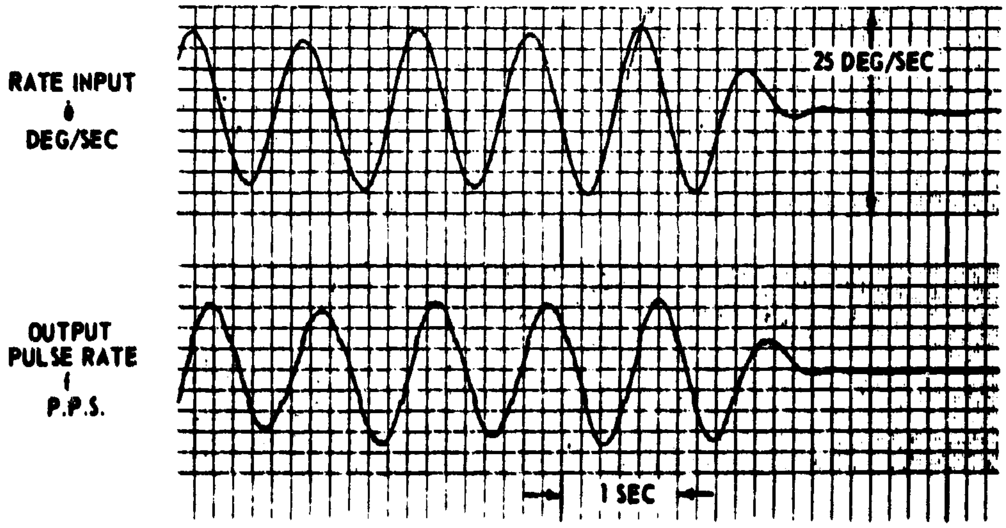


Fig. 10 Digital rate sensor performance

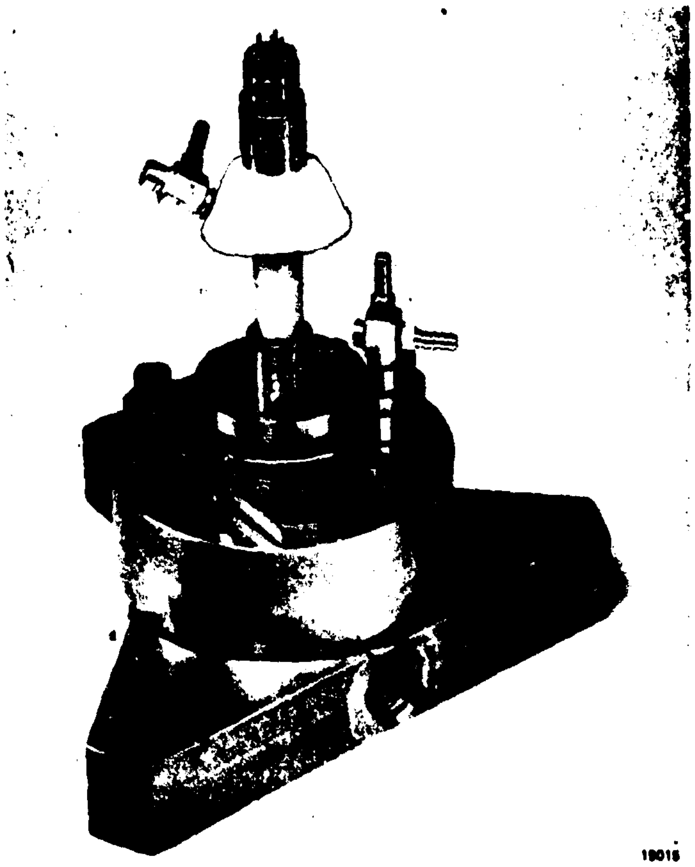
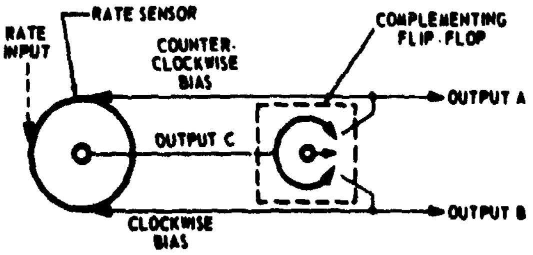
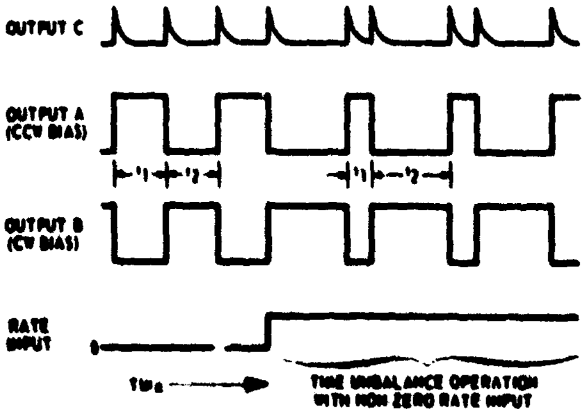


Fig. 11 Reversing bias rate sensor



(a) Schematic



(b) Output Wave Forms

Fig. 12 Basic rate sensor-signal processing sub-system

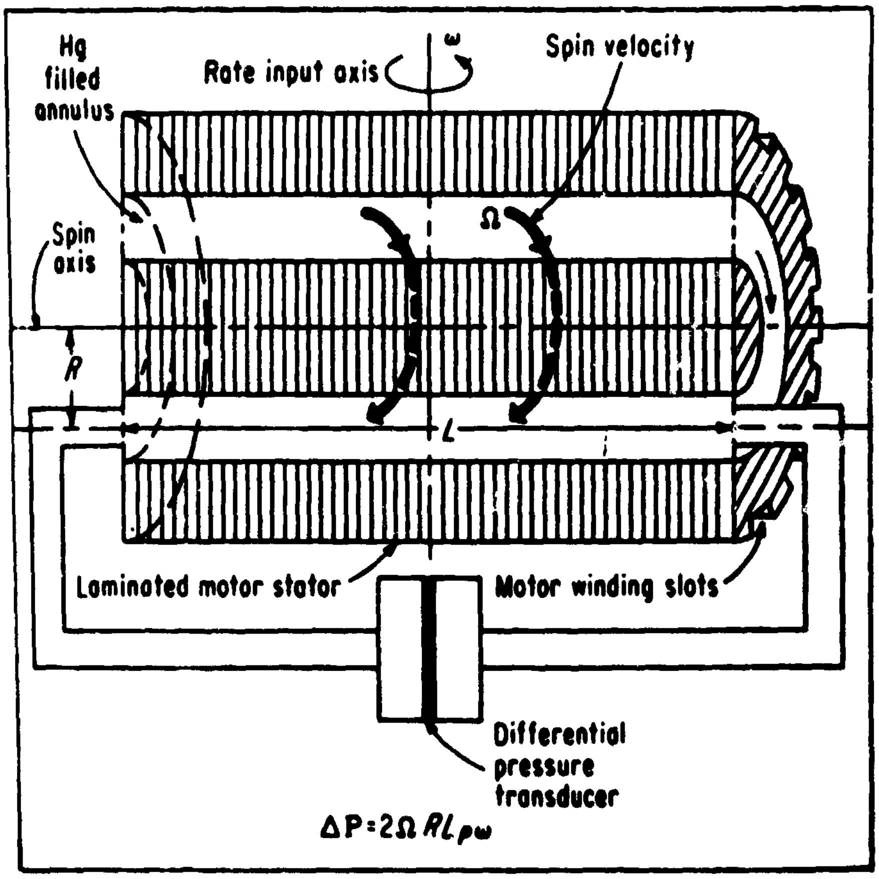


Fig. 13 Gyros with non-rotating containers

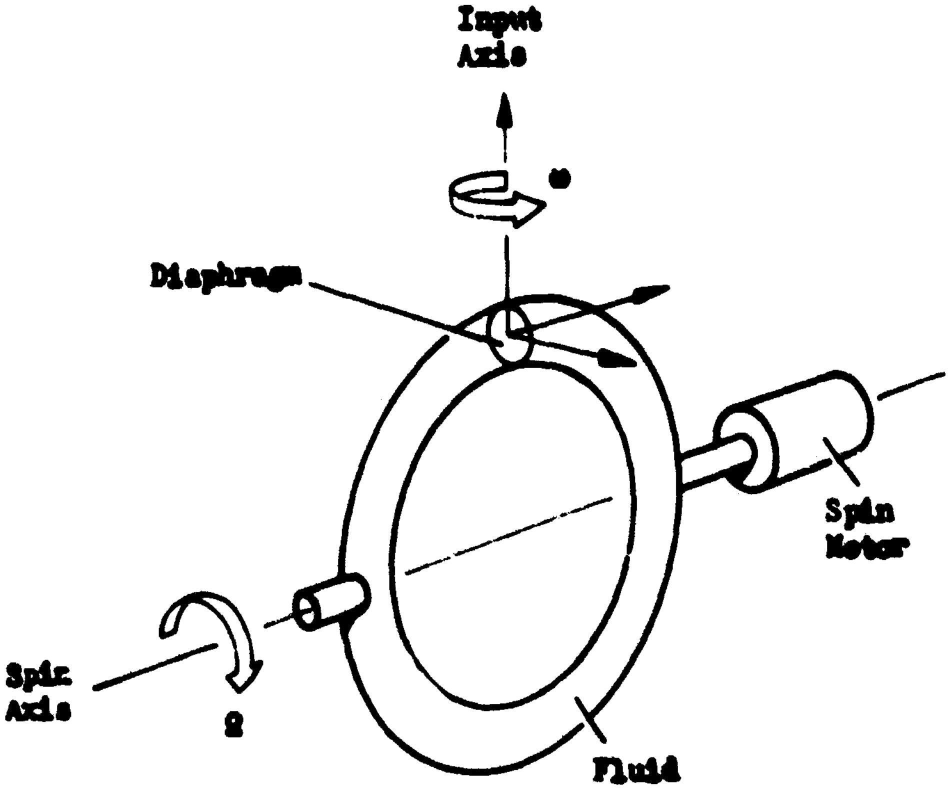


Fig. 14 Fluid tube gyro schematic

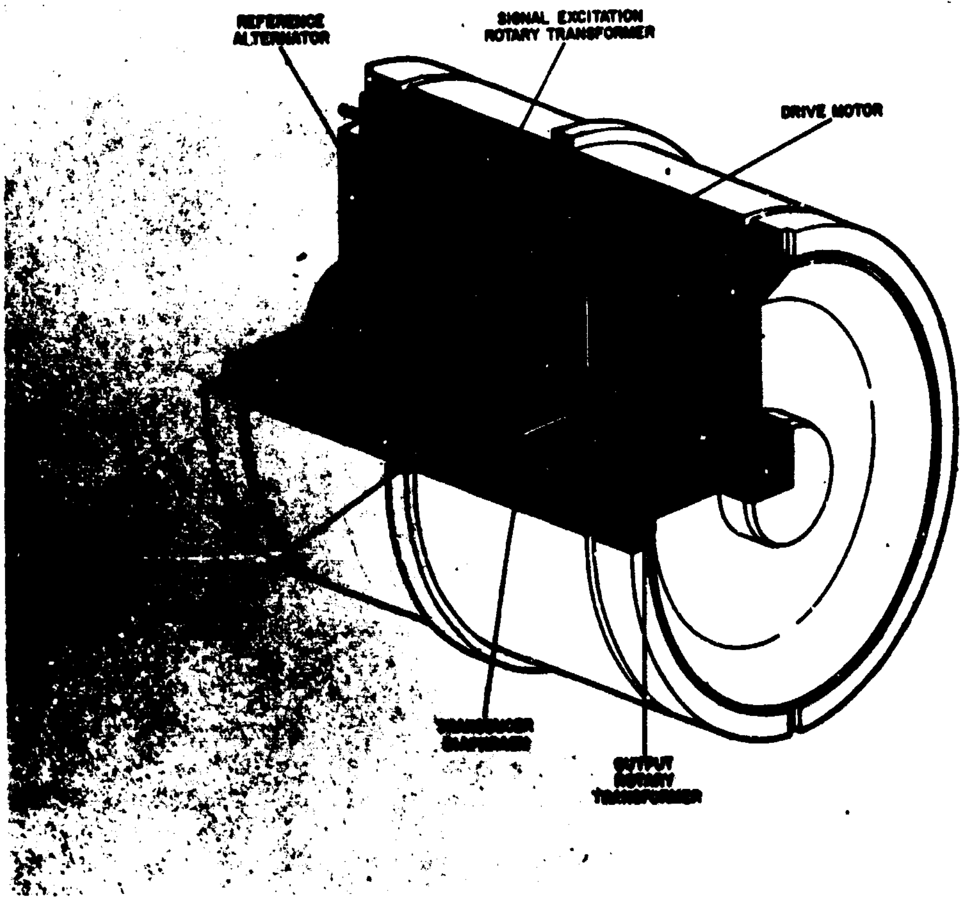

REFERENCE ALTERNATOR
SIGNAL EXCITATION ROTARY TRANSFORMER
DRIVE MOTOR
TRANSDUCER DIAPHRAGM
OUTPUT ROTARY TRANSFORMER

Figure 15

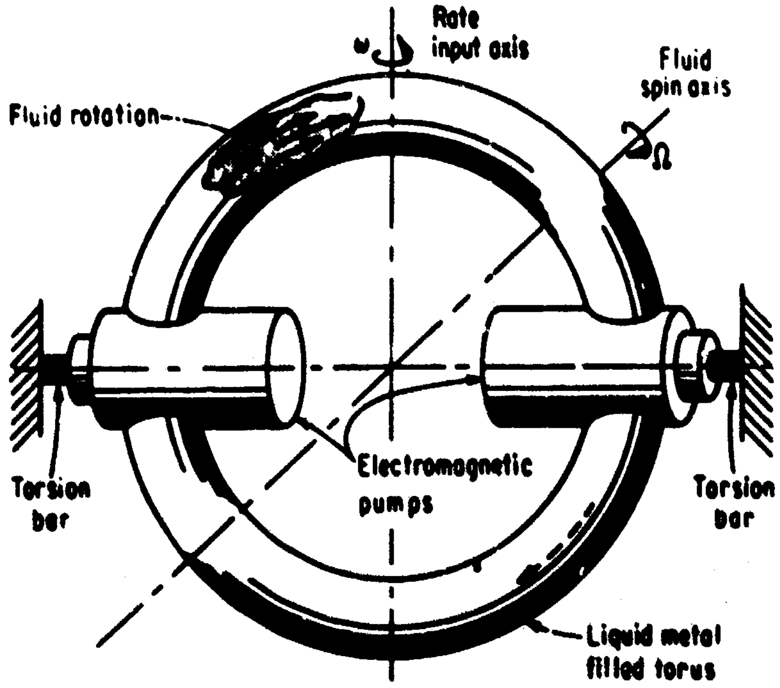


Fig. 16 Pumping fluid around spring-restrained torus produces fluid rotor gyro similar to conventional rate gyro. Rate sensing is mechanical angle measurement

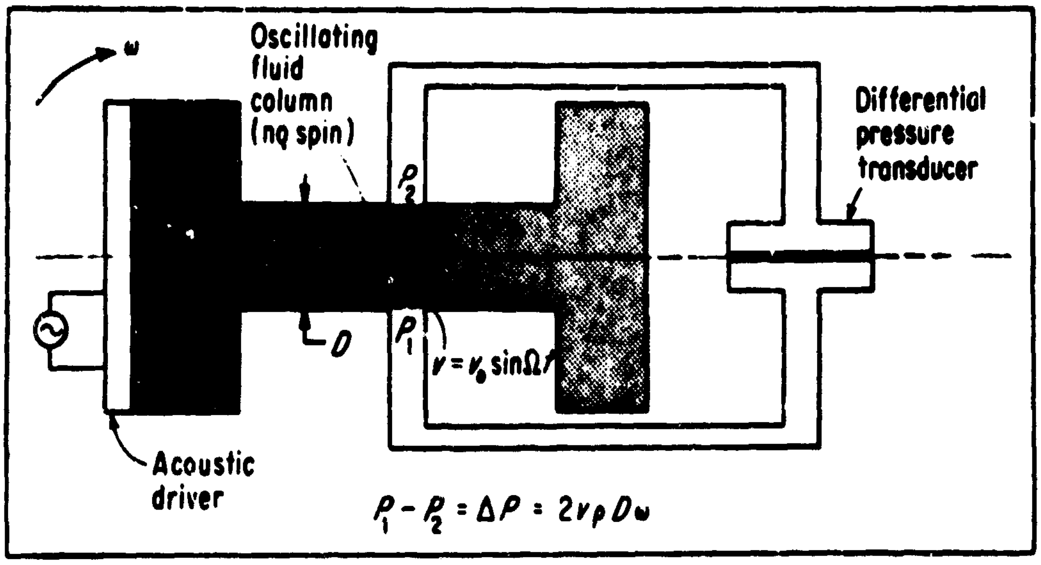

ω
Oscillating fluid column (no spin)
P_2
P_1
D
$v = v_0 \sin\Omega t$
Differential pressure transducer
Acoustic driver
$P_1 - P_2 = \Delta P = 2v\rho D\omega$

Fig. 17 Vibratory fluid column gyro is related to version of gyro in which mercury in annulus is caused to change spin direction at frequency Ω

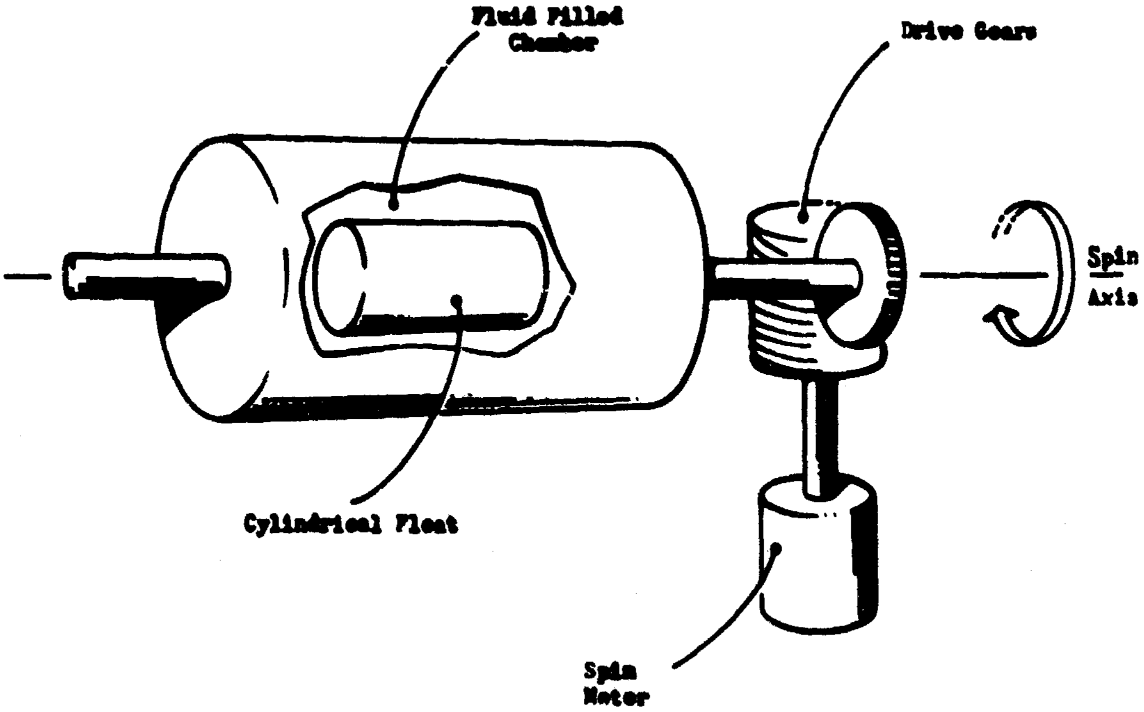


Fig. 18 Horizontal integrating accelerometer

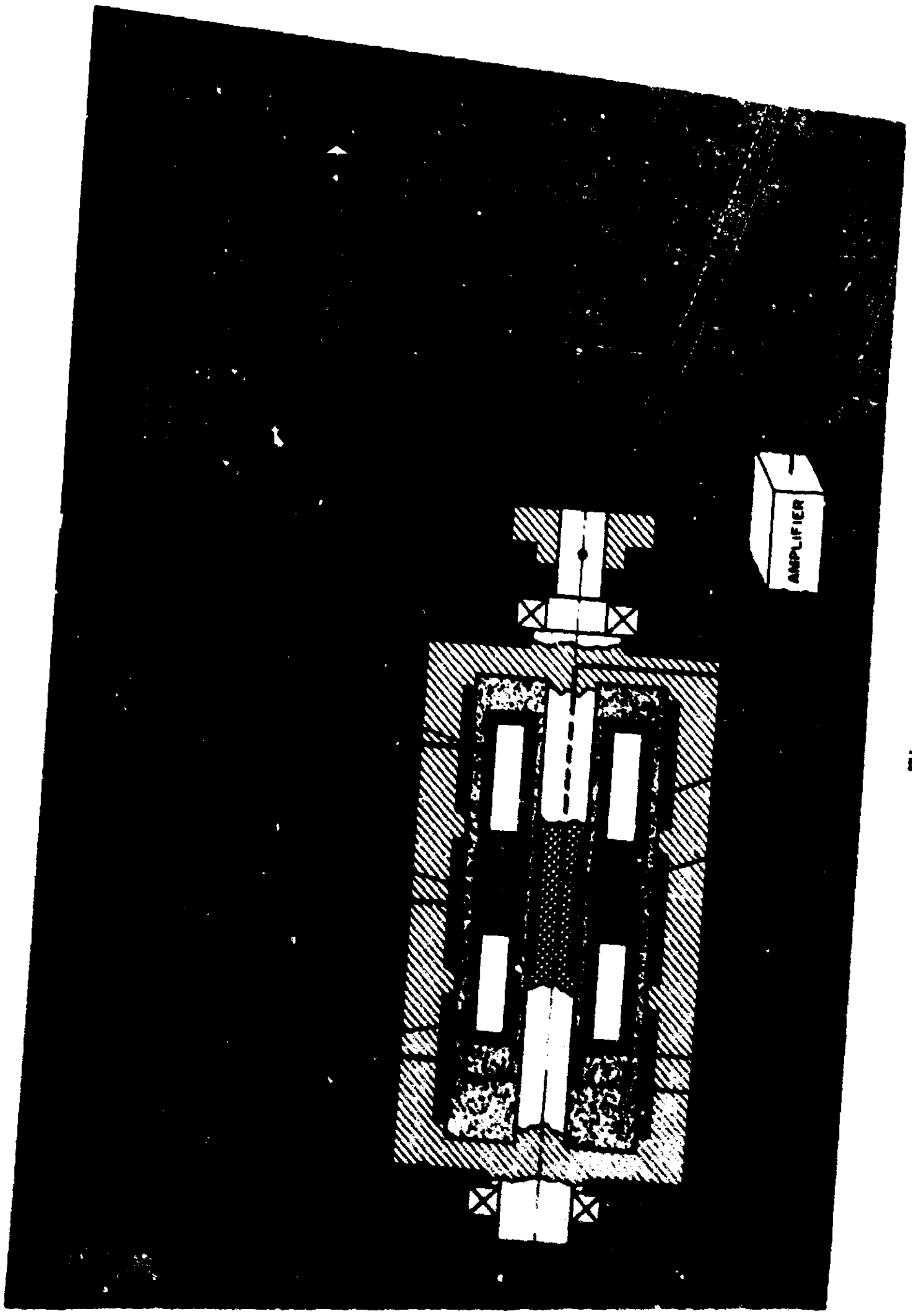


Figure 19

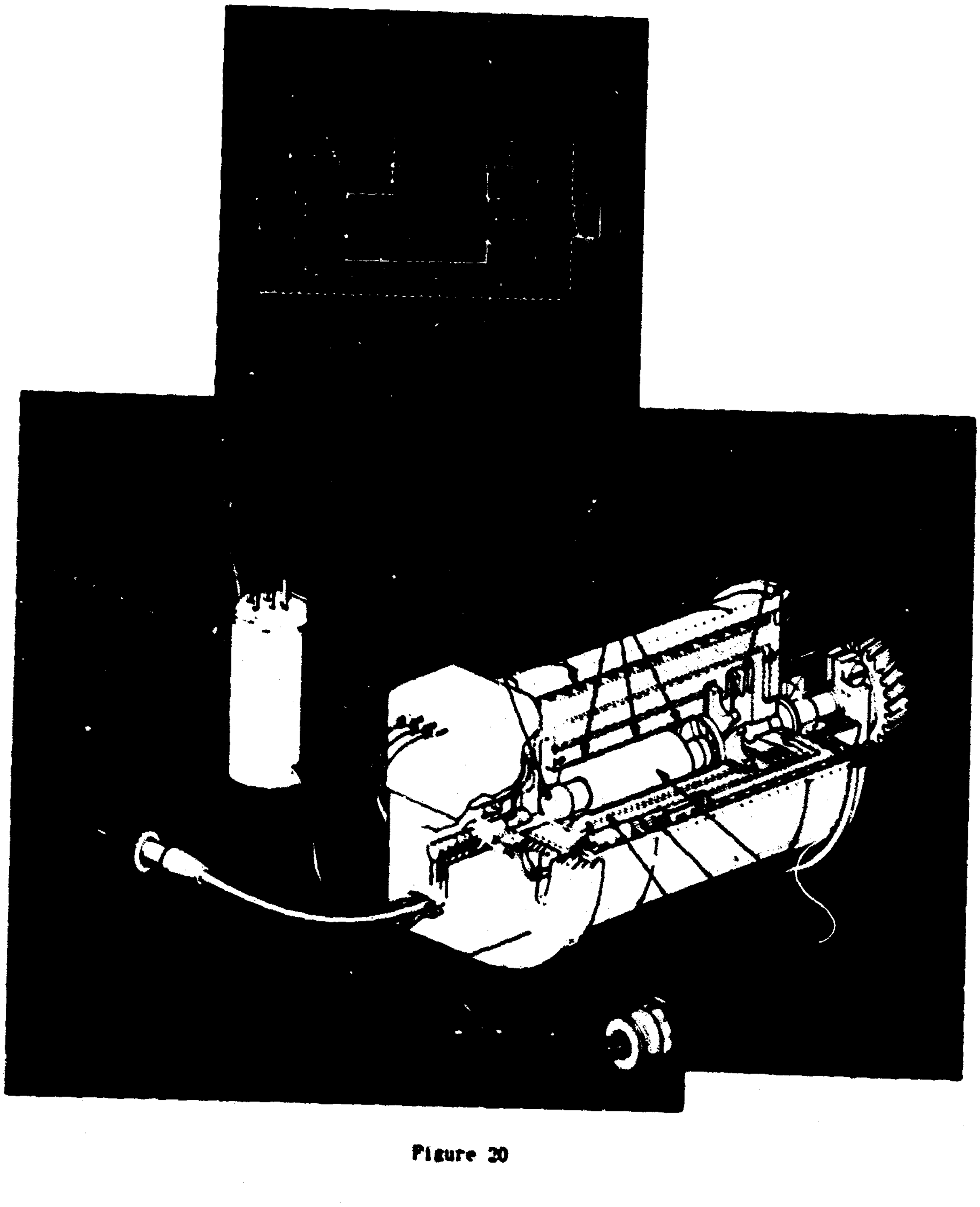


Figure 20

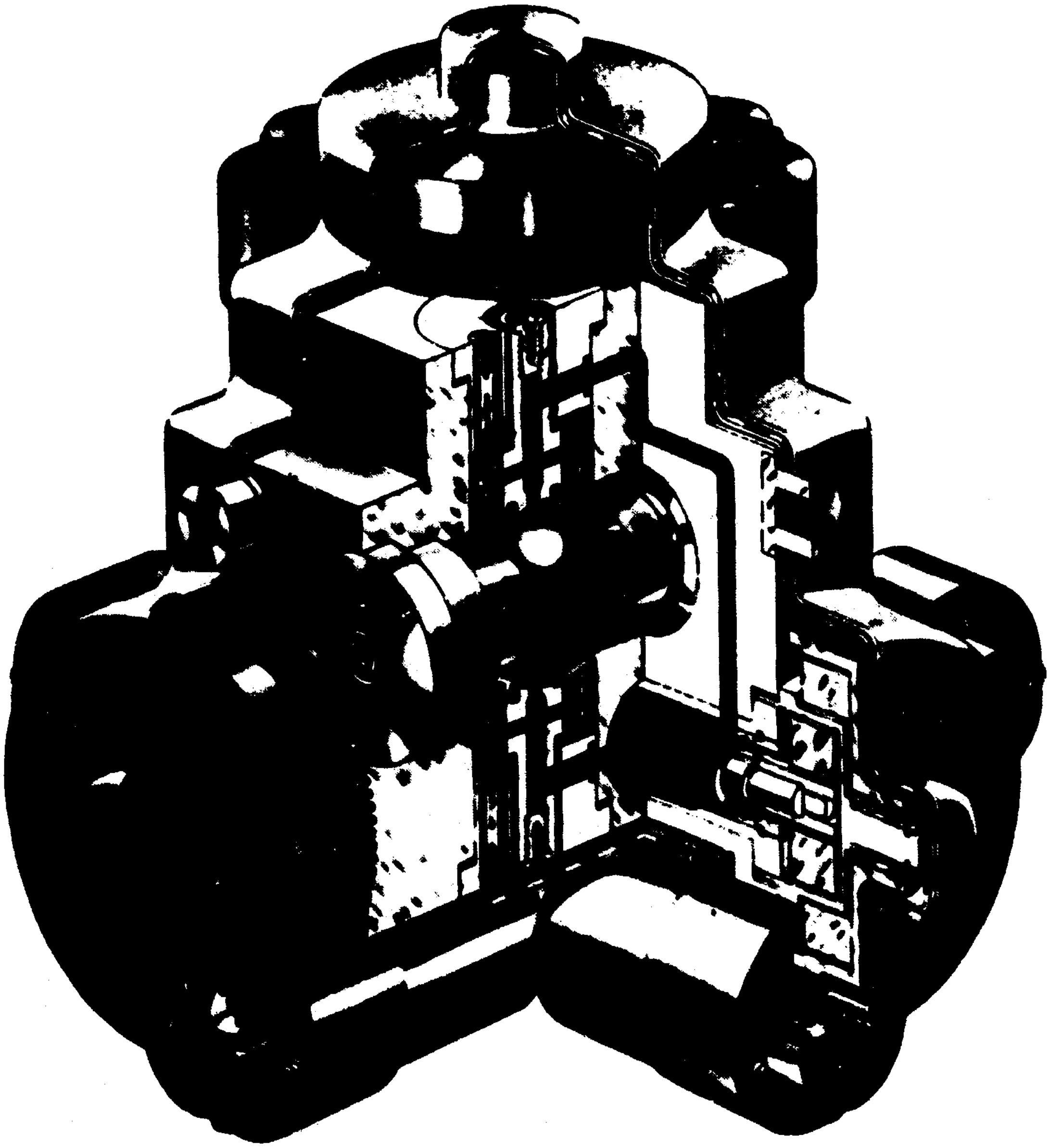


Figure 21

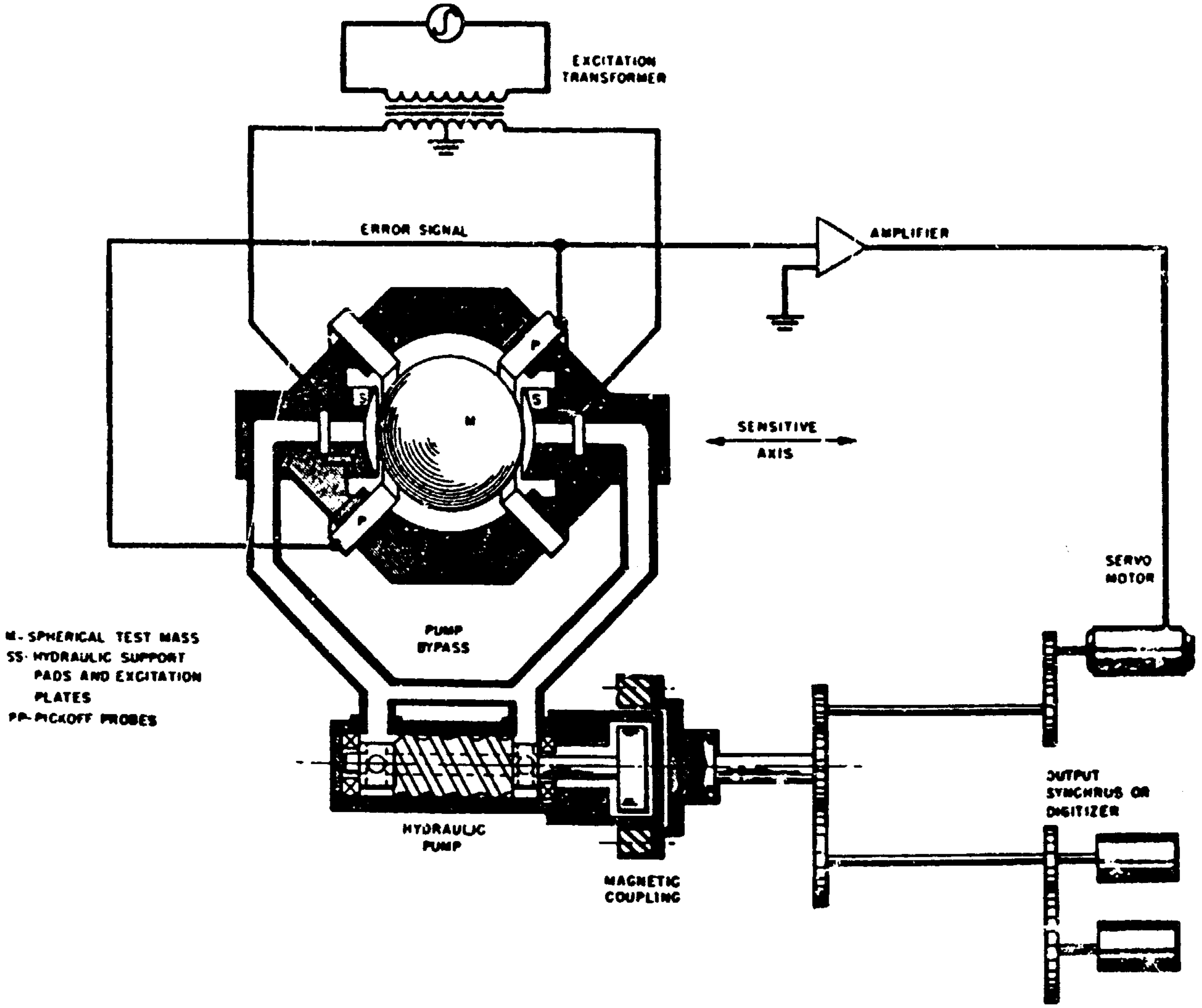


Fig. 22 One axis of a three-axis accelerometer, schematic diagram

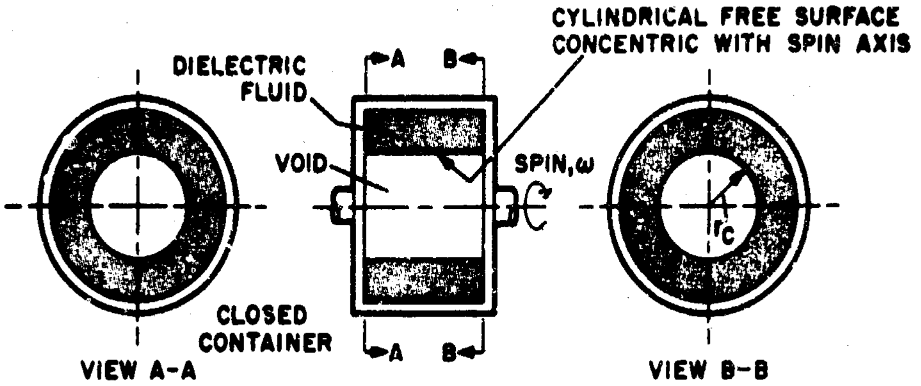


FIGURE A ZERO ACCELERATION

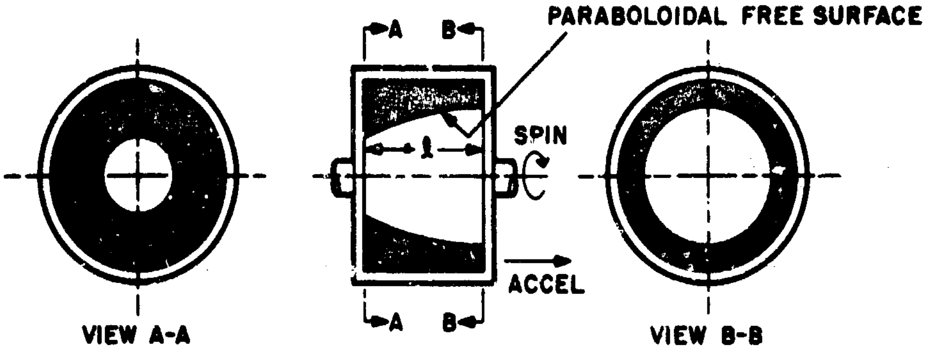


FIGURE B ACCELERATION TO THE RIGHT ALONG SPIN AXIS

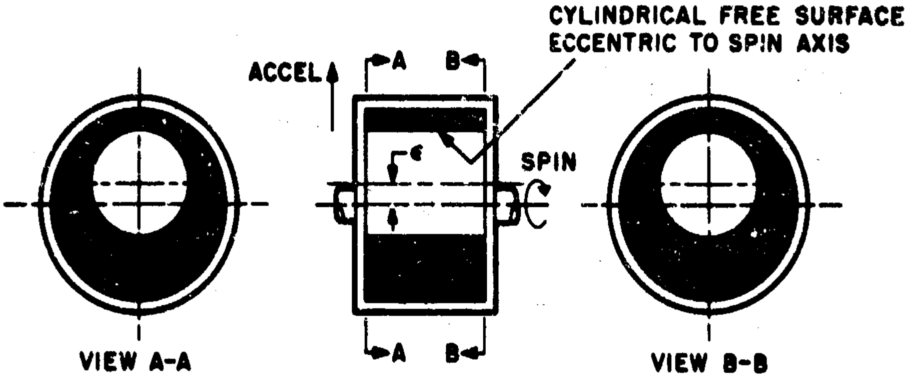


FIGURE C ACCELERATION UPWARD, NORMAL TO SPIN AXIS

Figure 23

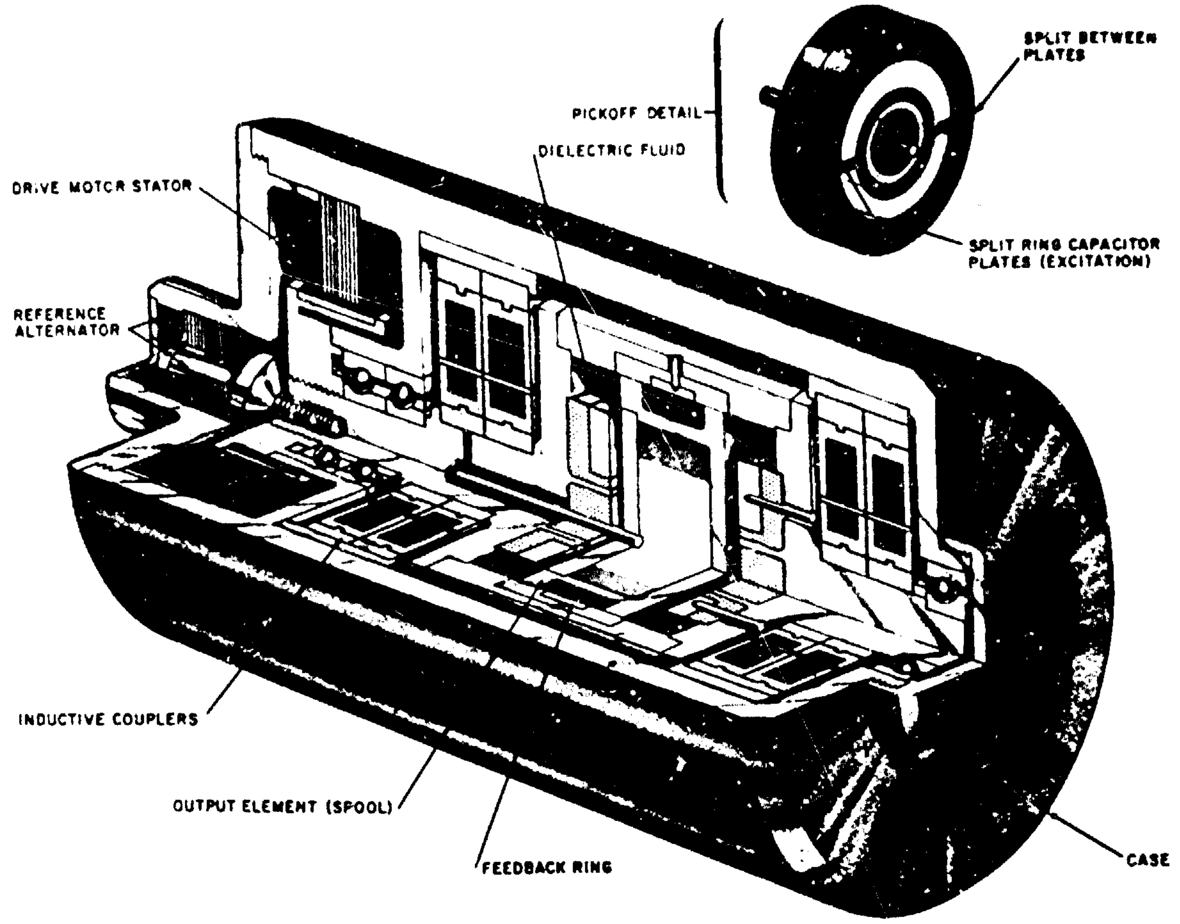

SPLIT BETWEEN PLATES
PICKOFF DETAIL
DIELECTRIC FLUID
DRIVE MOTCR STATOR
SPLIT RING CAPACITOR PLATES (EXCITATION)
REFERENCE ALTERNATOR
INDUCTIVE COUPLERS
OUTPUT ELEMENT (SPOOL)
FEEDBACK RING
CASE

Figure 24

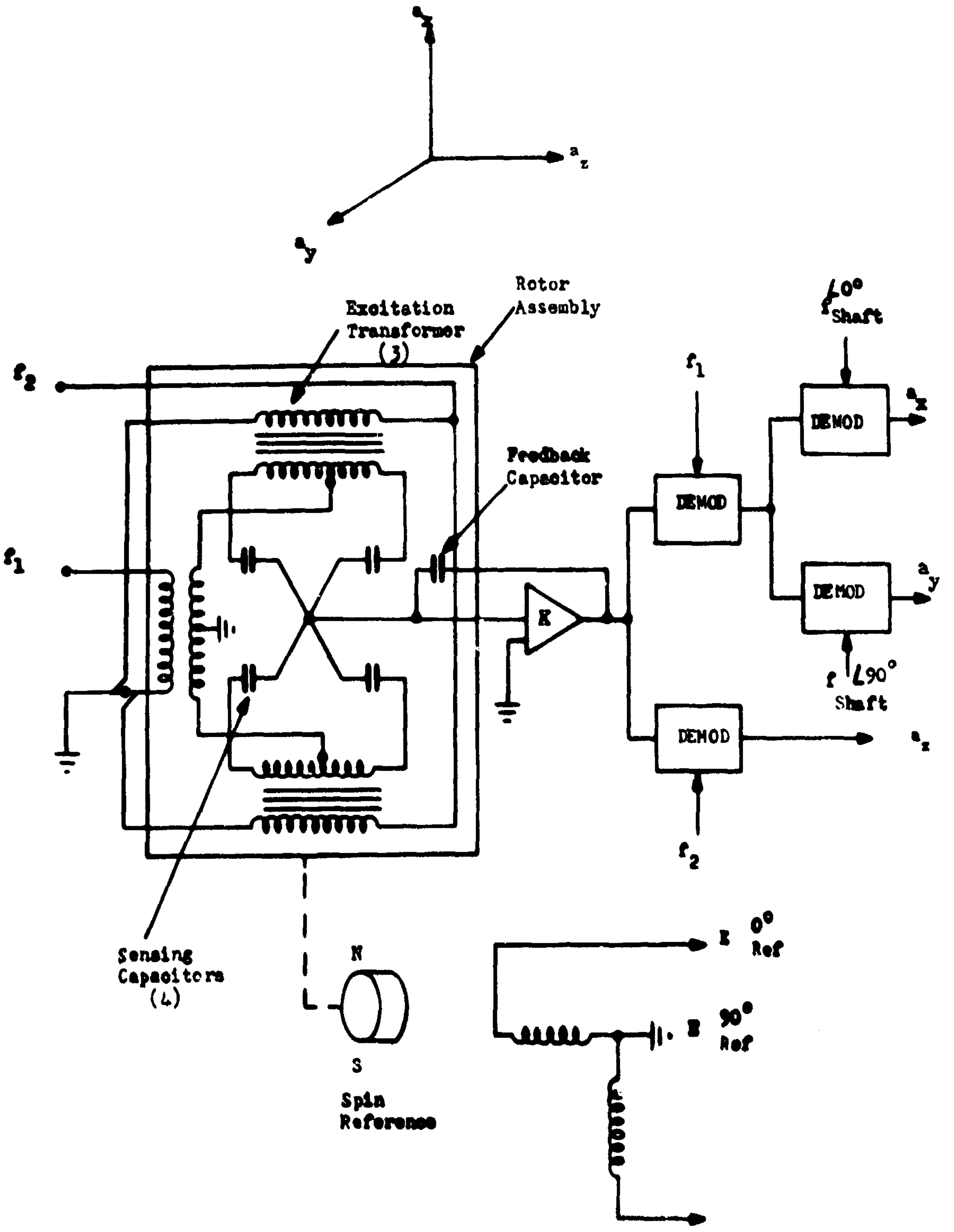


Fig. 25 Free-surface accelerometer. Signal processing

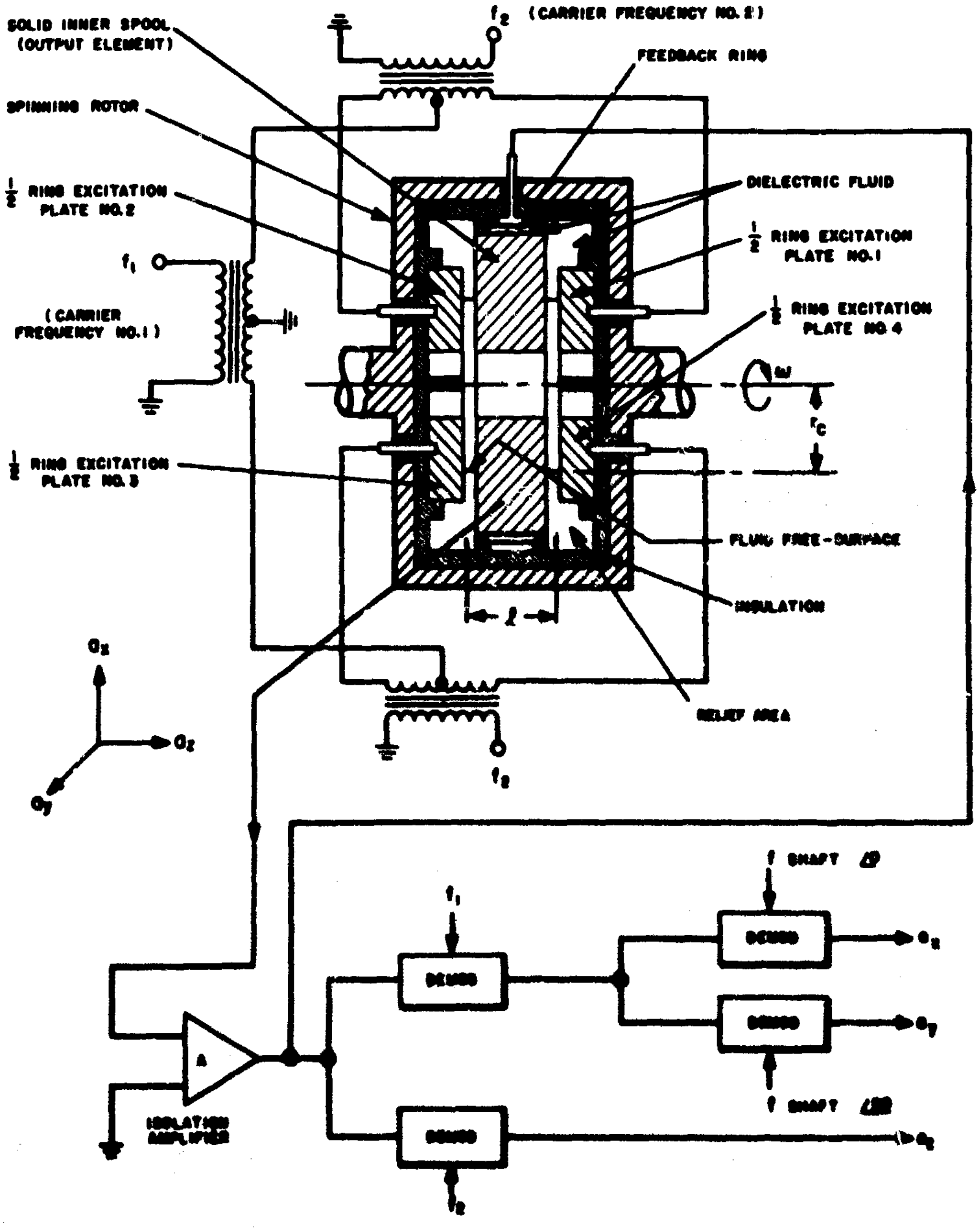

SOLID INNER SPOOL
(OUTPUT ELEMENT)
f_2 (CARRIER FREQUENCY NO. 2)
FEEDBACK RING
SPINNING ROTOR
$\frac{1}{2}$ RING EXCITATION PLATE NO. 2
DIELECTRIC FLUID
$\frac{1}{2}$ RING EXCITATION PLATE NO. 1
f_1
(CARRIER FREQUENCY NO. 1)
$\frac{1}{2}$ RING EXCITATION PLATE NO. 4
ω
r_c
$\frac{1}{2}$ RING EXCITATION PLATE NO. 3
FLUID FREE-SURFACE
INSULATION
ℓ
RELIEF AREA
Ω_x
Ω_z
Ω_y
f_2
f SHAFT $\angle 0$
f_1
DEMOD
Ω_x
DEMOD
Ω_y
f SHAFT $\angle 90$
A
ISOLATION AMPLIFIER
DEMOD
Ω_z
f_2

Figure 26